

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2001, 03:18:53 ; Search time 3860.4 Seconds
(without alignments)
8809.648 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtgggcctttgtctgct.....agaagtcagagccagtgta 2481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 seqs, 6853842396 residues 2236266
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

1: gb_bal:*
2: gb_bal:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pl1:*
7: gb_pl2:*
8: gb_pr1:*
9: gb_pr2:*
10: gb_pr3:*
11: gb_ro:*
12: gb_sy:*
13: gb_un:*
14: em_fun:*
15: em_hum1:*
16: em_hum2:*
17: em_in:*
18: em_om:*
19: em_or:*
20: em_ov:*
21: em_pat:*
22: em_ph:*
23: em_pl:*
24: em_ro:*
25: em_sts:*
26: em_sy:*
27: em_un:*
28: em_vi:*
29: gb_htg1:*
30: gb_htg2:*
31: gb_in1:*
32: gb_in2:*
33: gb_ba1:*
34: em_ba2:*
35: em_hum3:*
36: em_hum4:*
37: gb_pr4:*
38: gb_htg3:*
39: gb_htg4:*
40: gb_htg5:*
41: gb_htg6:*
42: gb_htg7:*
43: em_htg1:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pl3:*
48: gb_pr5:*
49: gb_htg8:*
50: gb_htg9:*
51: gb_htg10:*
52: gb_htg11:*
53: gb_htg12:*
54: gb_htg13:*
55: gb_htg14:*
56: gb_in3:*
57: gb_htg15:*
58: gb_htg16:*
59: gb_htg17:*
60: em_htg4:*
61: em_htg5:*
62: em_htg6:*
63: em_htg7:*
64: em_hum6:*
65: gb_htg18:*
66: gb_htg19:*
67: gb_htg20:*
68: gb_htg21:*
69: gb_htg22:*
70: gb_htg23:*
71: gb_v11:*
72: gb_v12:*
73: gb_ba3:*
74: em_htg8:*
75: em_htg9:*
76: em_htg10:*
77: gb_pr6:*
78: gb_pr7:*
79: gb_sts1:*
80: gb_sts2:*
81: gb_pat1:*
82: gb_pat2:*
83: em_htg0:*
84: gb_htg24:*
85: gb_pr8:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2448.2	98.7	2976	37	AK001392	AK001392 Homo sapi
2	247.4	10.0	118788	8	AC005277	AC005277 Homo sapi
3	133.4	5.4	2690	32	AF215894	AF215894 Drosophil
4	133.4	5.4	115873	41	AC017383	AC017383 Drosophil
5	133.4	5.4	125448	29	AC007417	AC007417 Drosophil
6	133.4	5.4	190574	29	AC007352	AC007352 Drosophil
7	133.4	5.4	261846	31	AE003830	AE003830 Drosophil
8	84.2	3.4	31201	47	SPAC1D4	269239 S.pombe chr
9	71.8	2.9	46335	6	AC022354	AC022354 Arabidops
10	70	2.8	19158	6	AF188714	AF188714 Emericell
11	67	2.7	32329	47	SPBC3D6	295620 S.pombe chr
12	65	2.6	143209	7	AP001550	AP001550 Oryza sat
13	58.8	2.4	75289	6	AB023046	AB023046 Arabidops
14	58.4	2.4	3685	47	SCYKR079C	228304 S.cerevisia
15	56.8	2.3	23533	1	AE001039	AE001039 Archaeogl
16	53.8	2.2	100005	65	AC079030	AC079030 Homo sapi
17	53.8	2.2	145440	52	AC026119	AC026119 Homo sapi
18	53.8	2.2	180260	57	AC068797	AC068797 Homo sapi
19	52.4	2.1	12026	1	AE005109	AE005109 Halobacte
20	51.8	2.1	11836	1	AE000936	AE000936 Methanoba
21	51.4	2.1	729	78	HSY18896	Y18896 Homo sapien

Db 675 GAAAGGCTCTCAGAGGCTCAGTCCAGAGCGATCTTCAGACTCGAGTGAATCAAAAT 734
Qy 661 gagccacaccttccacatggtgtagccagagaaggggtcagggaacttctccctggtc 720
Db 735 GAGCACACCTTCACATGTTGTTAGCCAGAGAGAGGGGTTCAGGGACTCTTCCTGGTC 794
Qy 721 gtacgttctcatctgaagcttccatcaaaagagagaaacttcttggctcaaaagcaag 780
Db 795 GTAGCTTCACTGTAGCTTCACTTAAAGAGAGAGAACTTCTTGGTGTCTCAAGCAAG 854
Qy 781 gagatgggctccacgttgggacagctgcccacatcctcccatcattgctgctgctcaagac 840
Db 855 GAGATGGGCTCCACAGTTGGACAGCTGCCATCGCTCCCATCATTTGCTGCTCAAGGAC 914
Qy 841 gggaaaagcatcactcatgaaggagagagattttggctgaagagctggtactccctcca 900
Db 915 GGGAAAAGCATCACTATGAAGAGAGAGAGATTTTGGCTGAAGAGCTGTGTACTCTCTCCA 974
Qy 901 gatcctggtgctgttttggtagagaaatgtccagatgaagacttcaatcaaccatc 960
Db 975 GATCCTGGTGTCTTGTGGTGGTGAATGTCCAGATGAAGCTTTCATTCAACCCCATC 1034
Qy 961 tgtgagatgccacttccagagtlaccaagaaagcagatgcccccggtggccttgggtg 1020
Db 1035 TGTGAGATGCCACTTTCAGAGGTACCAAGAAAGCGAGATGCCCCCTGGCCTTTGGTG 1094
Qy 1021 gttcacatggccccagcatctgtgttgacagcaggtaccagcagtgatggagagg 1080
Db 1095 GTTCACATGGCCCCAGCATCTGTCTTGTGACAGCAGGTACCAAGCAGTGGATGGAGAGG 1154
Qy 1081 tttgggctgacacccagcaacttgctcctgaatgagaaactgtgctcagtttcaaaacctt 1140
Db 1155 TTTGGGCTGTACACCCAGCACTTGTCTGTGAATGAGAACTGTGCTCTCAGTTCAACAACCTT 1214
Qy 1141 cgcagcacagaattcaaaaccagctcaacctcaccaccgagcatcttccccctgctc 1200
Db 1215 CCGAGCCACACAGATTCAAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCTCTGCTC 1274
Qy 1201 accagtttccgctgaagaagaggggccccaccctcagtggtcccatggttcaagggtgaa 1260
Db 1275 ACCAGTTTCCGCTGT --- AAGGAGGGCCCCACCCTCAGTGTGCCATGGTTCAAGGTGAA 1331
Qy 1261 tgcctcctcaagtaccagctccgtcccaaggagggtgagtgccagagggatgcattattact 1320
Db 1332 TGCTCTCTCAAGTACCAAGCTCCGTCCTCCAGAGGAGTGGCAGAGGGATGCCATTATTACT 1391
Qy 1321 tgcaatcctgaggaatcattagttgagcgtgagcttcccaacttccacagaaagctg 1380
Db 1392 TGCAATCTGAGGAATTCATAGTTGAGCGCTGCAGCTTCCCACTTCCAGCAGAGCGTG 1451
Qy 1381 caggagtacaggaggagtgccagagagggccccagccccagcagagagaaagtcagttac 1440
Db 1452 CAGGAGTACAGGAGGTGCGCAGAGCGGCCGCCAGCCCGCAGCAGAGAAAGTCAAGTAC 1511
Qy 1441 ccagaaatcattctccttggaaaggggtcgtgcattcccgatcccgatgcgaattcgaatgtcagt 1500
Db 1512 CCAGAAATCATCTCTCTTGGAAACAGGGTCTGCCATCCGATCCGATGAAGATTGCAAAATGTCAGT 1571
Qy 1501 gccacactgtcaacataagcccgacacgtctctgtactgagactggtgtaggggca 1560
Db 1572 CCCACACTGTACACATTAAGCCCCGACACGCTCTGTCTACTTGGAGTGGTGGAGGGCACA 1631
Qy 1561 tttgggcagctgtgcgtcattacggagaccagggtggacagggctcctgggcaacctggct 1620
Db 1632 TTTGGCAGCTGTGCCGTCATTACGGAGACCAGGTGGAGAGGGTCTTGGCACCTTGCT 1691
Qy 1621 gctgtgttgggttccacactgacgcagatcaccacacgggcttggcgaagtatttggctg 1680
Db 1692 GCTGTGTGTGTCTCCACCTGCAGCGCATCACCACACGGGCTTGCCAAAGTATCTTGCTG 1751
Qy 1681 cagagagaacgccttggatctttggaaagccggttccaccttttgcctggtggttccc 1740
Db 1752 CAGAGAGAACGCGCTTGGCAATCTTTGGAAAGCGGCTTCAACCTTTTCTGCTGGTGTGCC 1811

Qy 1741 cccaaaccagctcaaaagcctggctccagcaggtaccacaaccaggtccagaggaggtcctgcac 1800
Db 1812 CCCAACCAAGCTCAAAAGCCTGGCTCCAGCAGTACTAAACCAAGTGCAGAGGTCCTGCAC 1871
Qy 1801 cacatcagatgattcctgcctcaaatgcttcagggaagggctgagatctccagttcctgca 1860
Db 1872 CACATCAGATGATTCTTCGCAAAATGCCCTTCAGGAAGGGCTGAGATCTCCAGTCTCGCA 1931
Qy 1861 gtggaagatgatcagttcgtgttgcaacatgattttggaagagtttcagaccctgt 1920
Db 1932 GTGGAAAGATTGATCAGTTCCGCTGTTGCGAACATGTGATTTGGAAGAGTTTCAGACCTGT 1991
Qy 1921 ctgtgctggcactgcaagcatgctgttggctgtgctgctggtgcacacctctgctggcgaaa 1980
Db 1992 CTGCTGCGGCACTGCAAGCATGCGTTTGGCTGCGCTGGTGCACACCTCTGCTGGTGGAAA 2051
Qy 1981 gtggtctattccgggggacacacatgcccctgcagaggtctggtccggatggggaagatgcc 2040
Db 2052 GTGGTCTATTCCGGGGGACACCATGCCCTCCGAGGCTCTGGTCCGATGGGGAAGATGCC 2111
Qy 2041 accctcctgtatcatgaaagccacctggaagatggtttggaagagaaagcagtggaag 2100
Db 2112 ACCCTCTGTATCATCAAGCCACCTCGAAAGATGTTTGGAAAGAGACAGTGGAAAAAG 2171
Qy 2101 acacacagcaaacactcccaagccatcagcgtggggatgctggaatgaaacgcggagttcatt 2160
Db 2172 ACACACAGCAACAACGTCCTCAAGCCATCAGCGTGGGATGCGGATGAACGCGGAGTTCATT 2231
Qy 2161 atgctgaaccacttcagcagcagctatgccaaaggtccccctcttcagcccccaacttcagc 2220
Db 2232 ATGCTGAACCACTTTCAGCCAGCGCTATGCCAAGGTCCCCCTCTTCAGCCCCAACATCAGC 2291
Qy 2221 gagaaagtggaggtgctcttgacacacatgaaagctcgtcttggagacttccacaatg 2280
Db 2292 GAGAAAGTGGAGTGGCTTTGACCACATGAAGCTCTGCTTGGAGACTTTCACCAAGATG 2351
Qy 2281 cccaagctgattccccccactgaaagccctgttctgctgacatcagagatgagagag 2340
Db 2352 CCCAAGCTGATTCCCCCACTGAAGGCCCTGTTTGTGCGACATCGAGGAGATGGAGGAG 2411
Qy 2341 cgcagggagaagcggggagctgcggcaggtgcggcctcctctccagggagctggca 2400
Db 2412 CGCAGGAGAGAGCGGGAGCTGCGGAGGTGCGGGGGGGCCCTCTCTGTCCAGGGAGTGGCA 2471
Qy 2401 ggcgccttgagagatgggagcctcagcagaagcggccccacacagagagcagagcc 2460
Db 2472 GCGGCCCTGGAGGAT-GGGAGCCTCAGCAGAAGCGGGCCCCACACAGAGGACACAGGCC 2530
Qy 2461 aagaaggtcagagccccagtg 2481
Db 2531 AAGAAGGTCAAGGCCCAGTGA 2551

RESULT 2

AC005277 118788 bp DNA PRI 23-JUL-1998
LOCUS Homo sapiens chromosome 17, clone hRPK.597_M_12, complete sequence.
DEFINITION AC005277
ACCESSION AC005277
VERSION AC005277.1 GI:3337311
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 118788)
AUTHORS Birren, B., Fasman, K., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 17, clone hRPK.597_M_12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 118788)
AUTHORS Birren, B., Fasman, K., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Boatman, C.,
Boutwell, C., Brown, A., Castle, A., Cerny, J., Cooke, P., Depayre, E.,


```

Db 1744 CTGGCTCTCCGCAAAATTAACCTTGGTT---GGAGTTCTCAATCGACAATAGAAACC 1800
Qy 1795 ctgcaccacatcaglatgctcctgcccacaaatgccttcaggagggcgctgagatctccagt 1854
Db 1801 GTAGAAGATCGGTATACCTCTAGTGGCAACGGTGAACCTCTTGTCTAGTCCCTGAGCGGT 1860
Qy 1855 cctgcagtggaagattgatcagtcgctgtgtggaacatgattgttggaaagagttcag 1914
Db 1861 GAACAAGTTGAACGCTAGGAATACGTCATATCCACCTGCTAGTTAGGCACTGCCCC 1920
Qy 1915 acctgtctgtgctgctgactgaacatgcgttttggctgtgctgtgctgctgctgctgctg 1974
Db 1921 AACCTCTTCGGGAATAAGCCTTAACCTTGCGC-----GCAAGACACAATAGCGAAACC 1971
Qy 1975 tggaaagtggtctattccggtgggacaccatgcctgctgagagctctggtccggtgaggaaa 2034
Db 1972 GTCAAGATCACTTACAGCGGGGACACTATGCCATGTCAGGATCTGATCGATCTGGGCCGC 2031
Qy 2035 gatccaccctcctgtatatacgaagccaccctggaagatggtttggaagaggaagcagt 2094
Db 2032 GACTCCACCGTCTTATTCACGAGCGCAATGGAGGACGACTTGGAGGAGGAGGCGCGG 2091
Qy 2095 gaaaagacacacagcaaacgtcccaagccatcagcgtggtgggagtgatgaacgcggag 2154
Db 2092 TTAAGACGACAGCACCGCTGTCAGAGCTATTTCAGCAGGGGGCGGAATATGAACGCCCGC 2151
Qy 2155 ttcattatgctgaaccacttcagcagcgtatgcgaaggtcccccctt---cagcccc 2211
Db 2152 CATAGATTCTCACCCACTTTTCGAGCGGTACGCTAAGTCCCGCGGATTACCGAGTGAT 2211
Qy 2212 aactcagcgagaagtggtggagtgctcttgaccacatgaagtgctgtcttggagacttt 2271
Db 2212 GAGGATATGCAACGAGTTCGAATAGCCTTTTGATATATGAAGTGACCATTTGAGGATCTG 2271
Qy 2272 ccaacaatgcccagctgattccccactgaagccctgtttgtgtgctgctgctgctgctgctg 2331
Db 2272 CAGCACTACCAACAGCTCTACCCCGCCCTTTTCGCGATGTACGGCGAGTACACGGGAAGA 2331
Qy 2332 atggaggagcagggagagaagcggagctgctgctgctgctgctgctgctgctgctgctgctg 2372
Db 2332 CTGGAGCAGCGGCTGTCAACGGGAGCTGAACAGGAGCG 2372

```

```

RESULT 4
AC017383 115873 bp DNA HTG 09-DEC-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
DEFINITION pieces.
ACCESSION AC017383
VERSION AC017383.1 GI:6553603
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 115873)
Adams M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10210513 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..115873
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 31582 a 25315 c 25855 g 33121 t

```

ORIGIN

Query Match 5.4%; Score 133.4; DB 41; Length 115873;
 Best Local Similarity 50.5%; Pred. No. 6.5e-20;
 Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;

```

Qy 1438 taccacgaatacatctctcttggaaacagggcttgcctaccctccgcatgaagattcgaatgctc 1497
Db 63659 TACC CGAAGATAATTTCTTGGGCACAGGCTCTCTGCAATCCCAATAAGACGCGNAACGTG 63718
Qy 1498 agtgcacacattgtcaacataagccccgacacgctctctgtacttggaactgtgtgagggc 1557
Db 63719 AGCTCCATTCATGATCAGGACTGCAATTCATGCTACGCTACGCTGTTGTTGGATTGTGGAAGA 63778
Qy 1558 acatttggcagcgtgcgcgtcattacggagaccag---gtggacagggctcctgggcacc 1614
Db 63779 ACTTATGCTCAAAATGTTACGACTATATGGACACGAGAAAGGCGCATTTGATTTCTCGCCAG 63838
Qy 1615 ctggctgctgtgttgtgtcccccacctgcacgagatcacacacacgggcttgcgaagtac 1674
Db 63839 CTACAGGCGAATTTATGTTATCCCATTTGTCATGCCGACCAACCATTTGGATTGGTCTG 63898
Qy 1675 ttgctgcagagagaacgcgcttggcatcttbtgggaagccgcttccaccccttctgtgtg 1734
Db 63899 CTCCGGGAAAG-----GAGGCAATTAACACCTTAGGGCAGACCCACTTATTCTG 63946
Qy 1735 gttcccccacacagctcaaaagcctgcctccagcagtaaccacacagtgccagaggtc 1794
Db 63947 CTGGCTCTCCGCCAAATTTGAACCTTTGGTT---GGAGTTCTCAATCGACAATAAGAACCC 64003
Qy 1795 ctgcaccacatcagtatgattcctgcacaaatgccttcaggagggcgctgagatctccagt 1854
Db 64004 GTAGAAGATCGTATACCTACTAGTGGCAACGCTGAACCTCTAGCTAGTCCCTGAGCGGT 64063
Qy 1855 cctgcagtggaagattgatcagttcgtgtgtgtggaacatgattgttggaaagagttcag 1914
Db 64064 GAACAAGTTGAACGCTCTAGGAATTCAGTCCATATTCACCTGCTAGTTAGGCATCTGCC 64123
Qy 1915 acctgtctgtgctgctgactgcaacatgccttggctgtgctgctgctgctgctgctgctgctg 1974
Db 64124 AACTCTTTCGGAATAAGCCTTAACCTTGCGC-----GCAAGACACAATAGCGAACCC 64174
Qy 1975 tggaaagtggtctattccggtgggacaccatgcctgctgagagctctgctgctgctgctgctg 2034
Db 64175 GTCAAGATCACTTACAGCGGCGACACTATGCCATGCTCAGGATCTGATCGATCTGGGCCGC 64234
Qy 2035 gatccaccctcctgtatatacgaagccaccctggaagatggtttggaagaggaagcagt 2094
Db 64235 GACTCCACCGTCTTATTCACGAGCGCAATGGAGGATGACTTGGAGGAGGAGGCGCGG 64294
Qy 2095 gaaaagacacacagcaaacgtcccaagccatcagcgttgggagtgatgaacgcggag 2154
Db 64295 TTAAGACGACAGCACCGCTGTCGAGCTATTTCAGCAGGGCGGGAATATGAACGCCCGC 64354
Qy 2155 ttcattatgctgaaccacttcagcagcgtatgcgaaggtcccccctt---cagcccc 2211
Db 64355 CATAGATTCTCTCCCACTTTTCGAGCGGTTCAGCTAAGTCCCGCGGATTACCGAGTGAT 64414
Qy 2212 aacttcagcgagaagtggtggttgcctttgaccacatgaagaggtctgtctgttggagacttt 2271
Db 64415 GAGGATATGCAACGAGTTGCAATAGCCTTTTGATATATGAAGTGACCGTTGAGGATCTG 64474
Qy 2272 ccaacaatgcccagctgattccccactgaagccctgttgcctgctgctgctgctgctgctgctg 2331
Db 64475 CAGCACTACCAACAGCTCTACCCCGCCCTTTTCGCGATGTACGGCGAGTACACGGGAAGA 64534
Qy 2332 atggaggagcagggagagaagcggagctgctgctgctgctgctgctgctgctgctgctgctg 2372
Db 64535 CTGGAGCAGCGGCTGTCAAGCGGAGCTGAACAGGAGCG 64575

```

RESULT 5

AC007417/c

LOCUS AC007417 125448 bp DNA HTG 02-AUG-1999
 DEFINITION Drosophila melanogaster chromosome 2 clone BACR48F07 (D625) RPCI-98
 48.F.7 map 47A-47B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
 ***, 94 unordered pieces.

ACCESSION

VERSION AC007417.3 GI:5670592

KEYWORDS

SOURCE HTG; HTGS_PHASE1.

ORGANISM

Drosophila melanogaster
 fruit fly.
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS 1 (bases 1 to 125448)
 Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

TITLE Sequencing of Drosophila melanogaster

JOURNAL

REFERENCE

AUTHORS Unpublished
 2 (bases 1 to 125448)
 Celisner, S.E., Agbayan, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
 Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E.,
 Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.

TITLE Direct Submission

JOURNAL Submitted (28-Apr-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA

COMMENT

On Aug 2, 1999 this sequence version replaced gi:5629944.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 94 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 969: contig of 969 bp in length
 * 970 1049: gap of unknown length
 * 1050 2078: contig of 1029 bp in length
 * 2079 2158: gap of unknown length
 * 2159 3249: contig of 1091 bp in length
 * 3250 3329: gap of unknown length
 * 3330 4547: contig of 1218 bp in length
 * 4548 4627: gap of unknown length
 * 4628 5262: contig of 635 bp in length
 * 5263 5342: gap of unknown length
 * 5343 6096: contig of 754 bp in length
 * 6097 6176: gap of unknown length
 * 6177 7486: contig of 1310 bp in length
 * 7487 7566: gap of unknown length
 * 7567 8160: contig of 594 bp in length
 * 8161 8240: gap of unknown length
 * 8241 9090: contig of 850 bp in length
 * 9091 9170: gap of unknown length
 * 9171 9503: contig of 733 bp in length
 * 9904 9983: gap of unknown length
 * 10564: contig of 581 bp in length

10644: gap of unknown length
 11655: contig of 1011 bp in length
 11735: gap of unknown length
 12539: contig of 724 bp in length
 12539: gap of unknown length
 13197: contig of 658 bp in length
 13277: gap of unknown length
 14314: contig of 1037 bp in length
 14394: gap of unknown length
 15048: contig of 654 bp in length
 15128: gap of unknown length
 15691: contig of 963 bp in length
 16171: gap of unknown length
 16817: contig of 646 bp in length
 16897: gap of unknown length
 17598: contig of 701 bp in length
 17678: gap of unknown length
 18750: contig of 1072 bp in length
 18830: gap of unknown length
 19376: contig of 546 bp in length
 19456: gap of unknown length
 20255: contig of 799 bp in length
 20335: gap of unknown length
 20557: contig of 622 bp in length
 21037: gap of unknown length
 22187: contig of 1150 bp in length
 22267: gap of unknown length
 22871: contig of 604 bp in length
 22951: gap of unknown length
 24044: contig of 1093 bp in length
 24124: gap of unknown length
 25189: contig of 1065 bp in length
 25269: gap of unknown length
 26217: contig of 948 bp in length
 26297: gap of unknown length
 27312: contig of 1015 bp in length
 27392: gap of unknown length
 28334: contig of 1542 bp in length
 29014: gap of unknown length
 29998: contig of 984 bp in length
 30078: gap of unknown length
 31456: contig of 1378 bp in length
 31536: gap of unknown length
 33242: contig of 1706 bp in length
 33322: gap of unknown length
 34819: contig of 1497 bp in length
 34899: gap of unknown length
 35765: contig of 866 bp in length
 35845: gap of unknown length
 37032: contig of 1187 bp in length
 37112: gap of unknown length
 38879: contig of 1767 bp in length
 38959: gap of unknown length
 39889: contig of 930 bp in length
 39969: gap of unknown length
 40967: contig of 998 bp in length
 41047: gap of unknown length
 42656: contig of 1609 bp in length
 42736: gap of unknown length
 44172: contig of 1436 bp in length
 44252: gap of unknown length
 45669: contig of 1417 bp in length
 45749: gap of unknown length
 47740: contig of 1991 bp in length
 47820: gap of unknown length
 48827: contig of 1007 bp in length
 48907: gap of unknown length
 49770: contig of 863 bp in length
 49850: gap of unknown length
 51951: contig of 2101 bp in length
 52031: gap of unknown length
 53654: contig of 1623 bp in length
 53734: gap of unknown length

```

* 53735 55071: contig of 1337 bp in length
* 55072 53151: gap of unknown length
* 55152 56770: contig of 1619 bp in length
* 56771 56850: gap of unknown length
* 56851 58743: contig of 1893 bp in length
* 58744 58823: gap of unknown length
* 58824 60291: contig of 1468 bp in length
* 60371: gap of unknown length
* 63759: contig of 3388 bp in length
* 63760 63839: gap of unknown length
* 63840 66476: contig of 2637 bp in length
* 66477 66556: gap of unknown length
* 66557 69322: contig of 2765 bp in length
* 69322 69401: gap of unknown length
* 69402 70806: contig of 1405 bp in length
* 70807 70886: gap of unknown length
* 70887 73742: contig of 2856 bp in length
* 73743 73822: gap of unknown length
* 73823 75073: contig of 1253 bp in length
* 75076 75155: gap of unknown length
* 75156 77979: contig of 2824 bp in length
* 77980 78059: gap of unknown length
* 81375: contig of 3316 bp in length
* 81376 81455: gap of unknown length
* 81456 84452: contig of 2997 bp in length
* 84453 84532: gap of unknown length
* 84533 87596: contig of 3064 bp in length
* 87597 87676: gap of unknown length
* 87677 102529: contig of 14853 bp in length
* 102530 102609: gap of unknown length
* 102610 103128: contig of 519 bp in length
* 103129 103208: gap of unknown length
* 103209 103812: contig of 604 bp in length
* 103813 103892: gap of unknown length
* 103893 104112: contig of 220 bp in length
* 104113 104192: gap of unknown length
* 104193 104767: contig of 575 bp in length
* 104768 104847: gap of unknown length
* 104848 105570: contig of 723 bp in length
* 105571 105650: gap of unknown length
* 105651 106151: contig of 501 bp in length
* 106152 106231: gap of unknown length
* 106232 106941: contig of 710 bp in length
* 106942 107021: gap of unknown length
* 107022 107705: contig of 684 bp in length
* 107706 107786: gap of unknown length
* 107786 108412: contig of 627 bp in length
* 108413 108492: gap of unknown length
* 108493 109243: contig of 757 bp in length
* 109250 109329: gap of unknown length
* 109330 110040: contig of 711 bp in length
* 110041 110120: gap of unknown length
* 110121 110727: contig of 607 bp in length
* 110728 110807: gap of unknown length
* 110808 111413: contig of 606 bp in length

Query Match
Best Local Similarity 5.4%; Score 133.4; DB 29; Length 125448;
Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;

Qy 1438 taccagaaatcatctctcttggaacaggggtgccatcccgatgaagattcgaaatgtc 1497
Db 83158 TACCCGAAGATAATTTTTCGGCAGAGGCTCTGCAATCCCAATAGACGCAAAACGTG 83099

Qy 1498 agtccacactgtcaataaagccgacacgtctctgactgagctggtgagggc 1557
Db 83098 AGCTCCATCTGATCAGACTGCAATGTATGCCCTACGGTGTGTGGATTTGGAGAGGA 83039

Qy 1558 acatttgagcgtgtgccttcattacgagaccag---gtggacaggtcctgtggcacc 1614
Db 83038 ACTTATGTCACAAATTGACGACTATATGGACACGAGAAAGGCGATTGATCTTCGCCAG 82979

Qy 1615 ctggctgtgtttgtgtccacctgcaacgagatcacacacgagcggcttgccaaagtatc 1674

```

```

Db 82978 CTACAGGCGATTATGTATCCCATTTGATGCCGACCACCATATGGATTGATTGCTGTG 82919
Qy 1675 ttgtgcagagagaacgcgccttgatctcttggaagacgcgttcacaccttctgtgtg 1734
Db 82918 CTCCGGGAAAG-----GAGGCAATTAAACCTAGGGCAGACCCACTTATTCTG 82871
Qy 1735 gttgcccccaaccagctcaaacgctggctccagcagtagtaccacaaccagtgccagaggtc 1794
Db 82870 CTGGCTCTCCGCAAAATTGAACCTTGGTT---GGAGTTCTACAATCGACAAATAGAAC 82814
Qy 1795 ctgcaccacatcagtatgatctcctgcgaatgccttcgaaggggctgaatctccagt 1854
Db 82813 GTAGAAGATGCGTATACTTCTAGTGGCAACGCTGAACCTAGCTAGTCCCTGAGCGGT 82754
Qy 1855 cctgcagtggaaagatgatcagttcgtgttggaacatgattgattggaagagttcag 1914
Db 82753 GAACAAGTTGAACGCTAGGAATTACGTCCATATCCACTAGCTAGCTAGGACTGCCCC 82694
Qy 1915 acctgtctgtgtgcgacactgcaagcatgctttggctgtgcgctgtgcaacacctctg 1974
Db 82693 AACCTTTTCGGAATAAGCCTAACTCTGGCG-----GCAAAAGCAATAAGCAACCC 82643
Qy 1975 tggaaagtgtctatttcoggggacacatgcctctgcgaggtctgtggtccgagtgga 2034
Db 82642 GTCAAGATCACTTACAGCGCGACACTATGCATGTCAAGATCTGATCGATCTGGCGCG 82583
Qy 2035 gatccacctctctgatacatgaagcaccctggaagatggtttggaagagaagcagtg 2094
Db 82582 GACTCCACCGTCTTATTTCAGAGCGCAATGGAGGATGACTTGGAGGAGGAGCGCGG 82523
Qy 2095 gaaagacacacagcacacacgtcccaagccatcagctggtggatgcgagtgaaacgcgag 2154
Db 82522 TTAAGAGCGCACAGCAGCGTCTCGCAGGCTATTACAGCGGCGGGAATATGAACCGCC 82463
Qy 2155 ttcattatgtgaaccacttcagcgcagcgtatgcacaggttccccctt---cagcccc 2211
Db 82462 CACAGATTCTCTCCCACTTTTCGAGCGTTACGCTAAGTCCCGCGATTACCGAGTGAT 82403
Qy 2212 aacttcagcagagaagtggtggttccttgacacatgaaggtctgtgtggagacttt 2271
Db 82402 GAGGATATGCAACGAGTTGCAATAGCCTTTGATAATATGAAGTACCGTTGAGGATCTG 82343
Qy 2272 ccaacaaatgcccaagctgattcccccaactgaaagccctgttctgtggtgcagatcgagag 2331
Db 82342 CAGCATTACACACAGCTCTACCCGCGCCTTTTCGGGATGTACGCGGAGTACACGGAAGAA 82283
Qy 2332 atggagagcgcagggagagaagcggagcgtgcggcaggtgcg 2372
Db 82282 CTGGAGCAGCGGCTGTCAAGCGGAGCTGAACAGGAGCG 82242

RESULT 6
AC007352 190574 bp DNA HTG 24-FEB-2000
LOCUS Drosophila melanogaster chromosome 2 clone BACR19J16 (D626) RPCI-98
DEFINITION 19, J.16 map 47A-47B strain y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 36 unordered pieces.
AC007352
VERSION AC007352.8 GI:7025687
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 190574)
Celinker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B.,

```

Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,
Woolley, P., Tang, S., Yee, M., Yu, C. and Rubin, G.M.
Sequencing of Drosophila melanogaster

TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 190574)
Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Pfeiffer, B., Poon, L., Sequiera, A., Sethi, H., Snir, E.,
Svirskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.

Direct Submission
Submitted (22-APR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 24, 2000 this sequence version replaced gi:7018750.

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 36 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 678: contig of 678 bp in length
* 679 758: gap of unknown length
* 759 1358: contig of 600 bp in length
* 1359 1438: gap of unknown length
* 1439 2373: contig of 935 bp in length
* 2374 2453: gap of unknown length
* 2454 2962: contig of 509 bp in length
* 2963 3042: gap of unknown length
* 3043 3889: contig of 847 bp in length
* 3890 3969: gap of unknown length
* 3970 5437: contig of 1468 bp in length
* 5438 5517: gap of unknown length
* 5518 6572: contig of 1055 bp in length
* 6573 6652: gap of unknown length
* 6653 8387: contig of 1735 bp in length
* 8388 8467: gap of unknown length
* 8468 9060: contig of 593 bp in length
* 9061 9140: gap of unknown length
* 9141 11167: contig of 2027 bp in length
* 11168 11247: gap of unknown length
* 11248 14518: contig of 3271 bp in length
* 14519 14598: gap of unknown length
* 14599 18937: contig of 4239 bp in length
* 18938 18917: gap of unknown length
* 18917 24324: contig of 5407 bp in length
* 24325 24404: gap of unknown length
* 24405 29984: contig of 5580 bp in length
* 29985 30064: gap of unknown length
* 30065 35794: contig of 5730 bp in length
* 35795 35874: gap of unknown length
* 35875 44808: contig of 8934 bp in length
* 44809 44888: gap of unknown length
* 44889 58222: contig of 13334 bp in length
* 58223 58302: gap of unknown length
* 58303 78322: contig of 20020 bp in length
* 78323 78402: gap of unknown length
* 78403 122249: contig of 43847 bp in length
* 122250 122329: gap of unknown length
* 122330 179102: contig of 56773 bp in length
* 179103 179182: gap of unknown length
* 179183 179407: contig of 225 bp in length
* 179408 179487: gap of unknown length

* 179488 180592: contig of 1105 bp in length
* 180593 180672: gap of unknown length
* 180673 181383: contig of 711 bp in length
* 181384 181463: gap of unknown length
* 181464 182092: contig of 629 bp in length
* 182093 182172: gap of unknown length
* 182173 183480: contig of 1308 bp in length
* 183481 183561: gap of unknown length
* 183561 183771: contig of 211 bp in length
* 183772 183851: gap of unknown length
* 183852 184058: contig of 207 bp in length
* 184059 184138: gap of unknown length
* 184139 185035: contig of 897 bp in length
* 185036 185115: gap of unknown length
* 185116 185846: contig of 731 bp in length
* 185847 185926: gap of unknown length
* 185927 186560: contig of 634 bp in length
* 186561 186640: gap of unknown length
* 186641 187533: contig of 613 bp in length
* 187254 187334: gap of unknown length
* 187334 187611: contig of 278 bp in length
* 187612 187691: gap of unknown length
* 187692 188409: contig of 718 bp in length
* 188410 188489: gap of unknown length
* 188490 189196: contig of 707 bp in length
* 189197 189276: gap of unknown length
* 189277 189969: contig of 693 bp in length
* 189970 190049: gap of unknown length
* 190050 190574: contig of 525 bp in length.

FEATURES
source

1..190574
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2"
/map="47A-47B"
/clone="BACR19J16 (D626) RPCI-98 19.J.16"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACs3.6"
BASE COUNT 53370 a 39726 c 40358 g 54299 t 2821 others
ORIGIN

Query Match 5.4%; Score 133.4; DB 29; Length 190574;
Best Local Similarity 50.5%; Pred. No. 6.2e-20;
Matches 475; Conservative 0; Mismatches 436; Indels 30; Gaps 5;

QY 1438 taaccagaaatcatcttccttggaaacagggtctgccatcccgatgaagattcgaatgtc 1497
DB 132812 TACCCGAAGATAATTTCTTGGGCACAGGCTCCTGCAATCCCAATAAGACGCGAAACGTG 132871
QY 1498 agtgcacacattgtcaacataaagccccgcagacgtctctgtactgactgtgtgagggc 1557
DB 132872 AGCTCCATTCTGATCAGGACTGCAATTGATGCTTGTGATTTGTGGAGAGGA 132931
QY 1558 acattgggcagctgtgcgtcattacggagaccag---gtgcacaggggtccctgggcacc 1614
DB 132932 ACTTATGGTCAAAATTTGATACGACTATATGGACACCAAGGCGCATTTGATTTCCCGAC 132991
QY 1615 ctggctgctgtgttcttccacacctgcacgcagatcacacacagcggttgcgaagtac 1674
DB 132992 CTACAGCGCATTTATGTATCCCATTTGATTCGCGACCAACACACATTTGATTTGGTCTG 133051
QY 1675 ttgctgcagagagaaacgccttggcatttttgggaaagcgcgttccaccttgcctggtg 1734
DB 133052 CTCCTGGGAAAG-----GAGGCAATTAACCTAGGCGACAGCCACCATTTATCTG 133099
QY 1735 gttgcccccaacacagctcaaacgctgctccagcagatcacacacacagctgcccagggc 1794
DB 133100 CTGGCTCCCTGGCCAAATTTGAACCTTGGTT---GGAGTTTACAAATGACAAATAGAAAC 133156
QY 1795 ctgcaccacatcagatgattctctgccttcacaaatgccttcacgaagggtgctgagatctccagt 1854

after the database hits. Genes without significant peptide similarity but with ESI similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Location/Qualifiers

FEATURES

source	1. .46335 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="1" /map="m213" /clone="F915" 1. .5778 /note="overlap with BAC clone F5f19 (AC006216:105116. .110893)."	repeat_region	13177. .13298 /note="novel repeat from T25M19 22823 to 23610 788 nt"
misc_feature	1. .5778 /note="overlap with BAC clone F5f19 (AC006216:105116. .110893)."	repeat_region	15264. .15339 /rpt_family="(GAAA)n" complement(18843. .18918) /rpt_family="(GAAA)n" 19668. .19760 /rpt_family="POLY_A" complement(20890. .20985) /rpt_family="(GA)n" complement(20931. .20985) /rpt_family="(GAA)n" join(<21044. .21140,21233. .21450,23180. .23739,24098. .25199) /gene="F915.4" complement(<21044. .25199 /gene="F915.4" /note="similar to PEPTIDE TRANSPORTER PTR2-B GB:P46032 from [Arabidopsis thaliana]" join(21044. .21140,21233. .21450,23180. .23739,24098. .25046) /gene="F915.4" /codon_start=1 /product="putative peptide transporter" /protein_id="AAF29404.1" /db_xref="GI:6850341" /translation="MENPNETEAKIQITNEGKTKGGIITMPFIIANEAFKVASVG LLPNMILYLRDYPFGVAKGTNVLPMASASNTPLLGAFLSDSYLGRLFTISLASL SFLGMVLLMTAMLPQVRKSPCDTPAAGSHCGSTASQALYSAFALISIGSGIRP CSLAFGADQDNKNERNVLEFSGYASSAVAVIAFTGIVYIOEHLGKIGFG VPALMLIAALLFLASPLYVTVGTGKSLFTGLAQATVAAYKKRKLSPDHPHSDFCY YHMKSEIKAPSKOKLRLNKAACLIISNREEIGSDGFPALNPWLCTTDKVELKALIKV IPIWSTGIMNSTSSQQLQATSMRRLSRHSGSFQVPGAGFMTIILALAWI LDRAVIPASKIRGPRFLSVKRLMGLFMSFLAMASVMVESFRKKAISOQYAN NSNAVVDISAMMLVPOYVLGLAEALTAIGQTEFFTEPKSMSSAASLGLGMAVA SLASVLANAVNELTSRNGKESWVSDNINKHYNYIYVWLAINMSFINIYVICWSY GLPVDVRNGVNGVREELIDIVGKFEKEDLSPPVKTN" 26156. .26179 /rpt_family="(TA)n" 26779. .26899 /rpt_family="(CAAAA)n" join(<26899. .26982,27574. .>27987) /gene="F915.5" complement(<26899. .>27987 /gene="F915.5" /note="predicted by genscan" join(26899. .26982,27574. .27987) /gene="F915.5" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29405.1" /db_xref="GI:6850342" /translation="MMRYCQTPDYPSSNMDDLKAFEDETWYKGGKKEAWSLFDVLY SEEMKNMGEYTLINMYNECFKIGRFDQAIOFYKTKALKLNPDVTCYRNITTLNKO GVLFEEAYFEEMCSDRLVPPDVFTYTTMIDAYLKAGKTEDALRISKNMVDALFGQVA WLACL" complement(27450. .27491) /rpt_family="(TA)n" complement(join(<29019. .29311,29402. .29473,29582. .>29780)) /gene="F915.6" complement(<29019. .>29780) /gene="F915.6" /note="similar to PGP224 GB:AAD02554 from [Petunia x hybrida]" complement(join(29019. .29311,29402. .29473,29582. .29780))
misc_feature	1 (AC006216:105116. .110893)."	repeat_region	19668. .19760 /rpt_family="POLY_A" complement(20890. .20985) /rpt_family="(GA)n" complement(20931. .20985) /rpt_family="(GAA)n" join(<21044. .21140,21233. .21450,23180. .23739,24098. .25199) /gene="F915.4" complement(<21044. .25199 /gene="F915.4" /note="similar to PEPTIDE TRANSPORTER PTR2-B GB:P46032 from [Arabidopsis thaliana]" join(21044. .21140,21233. .21450,23180. .23739,24098. .25046) /gene="F915.4" /codon_start=1 /product="putative peptide transporter" /protein_id="AAF29404.1" /db_xref="GI:6850341" /translation="MENPNETEAKIQITNEGKTKGGIITMPFIIANEAFKVASVG LLPNMILYLRDYPFGVAKGTNVLPMASASNTPLLGAFLSDSYLGRLFTISLASL SFLGMVLLMTAMLPQVRKSPCDTPAAGSHCGSTASQALYSAFALISIGSGIRP CSLAFGADQDNKNERNVLEFSGYASSAVAVIAFTGIVYIOEHLGKIGFG VPALMLIAALLFLASPLYVTVGTGKSLFTGLAQATVAAYKKRKLSPDHPHSDFCY YHMKSEIKAPSKOKLRLNKAACLIISNREEIGSDGFPALNPWLCTTDKVELKALIKV IPIWSTGIMNSTSSQQLQATSMRRLSRHSGSFQVPGAGFMTIILALAWI LDRAVIPASKIRGPRFLSVKRLMGLFMSFLAMASVMVESFRKKAISOQYAN NSNAVVDISAMMLVPOYVLGLAEALTAIGQTEFFTEPKSMSSAASLGLGMAVA SLASVLANAVNELTSRNGKESWVSDNINKHYNYIYVWLAINMSFINIYVICWSY GLPVDVRNGVNGVREELIDIVGKFEKEDLSPPVKTN" 26156. .26179 /rpt_family="(TA)n" 26779. .26899 /rpt_family="(CAAAA)n" join(<26899. .26982,27574. .>27987) /gene="F915.5" complement(<26899. .>27987 /gene="F915.5" /note="predicted by genscan" join(26899. .26982,27574. .27987) /gene="F915.5" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29405.1" /db_xref="GI:6850342" /translation="MMRYCQTPDYPSSNMDDLKAFEDETWYKGGKKEAWSLFDVLY SEEMKNMGEYTLINMYNECFKIGRFDQAIOFYKTKALKLNPDVTCYRNITTLNKO GVLFEEAYFEEMCSDRLVPPDVFTYTTMIDAYLKAGKTEDALRISKNMVDALFGQVA WLACL" complement(27450. .27491) /rpt_family="(TA)n" complement(join(<29019. .29311,29402. .29473,29582. .>29780)) /gene="F915.6" complement(<29019. .>29780) /gene="F915.6" /note="similar to PGP224 GB:AAD02554 from [Petunia x hybrida]" complement(join(29019. .29311,29402. .29473,29582. .29780))
repeat_region	submitted sequence due to an overlap with another BAC (F5f19)."	repeat_region	19668. .19760 /rpt_family="POLY_A" complement(20890. .20985) /rpt_family="(GA)n" complement(20931. .20985) /rpt_family="(GAA)n" join(<21044. .21140,21233. .21450,23180. .23739,24098. .25199) /gene="F915.4" complement(<21044. .25199 /gene="F915.4" /note="similar to PEPTIDE TRANSPORTER PTR2-B GB:P46032 from [Arabidopsis thaliana]" join(21044. .21140,21233. .21450,23180. .23739,24098. .25046) /gene="F915.4" /codon_start=1 /product="putative peptide transporter" /protein_id="AAF29404.1" /db_xref="GI:6850341" /translation="MENPNETEAKIQITNEGKTKGGIITMPFIIANEAFKVASVG LLPNMILYLRDYPFGVAKGTNVLPMASASNTPLLGAFLSDSYLGRLFTISLASL SFLGMVLLMTAMLPQVRKSPCDTPAAGSHCGSTASQALYSAFALISIGSGIRP CSLAFGADQDNKNERNVLEFSGYASSAVAVIAFTGIVYIOEHLGKIGFG VPALMLIAALLFLASPLYVTVGTGKSLFTGLAQATVAAYKKRKLSPDHPHSDFCY YHMKSEIKAPSKOKLRLNKAACLIISNREEIGSDGFPALNPWLCTTDKVELKALIKV IPIWSTGIMNSTSSQQLQATSMRRLSRHSGSFQVPGAGFMTIILALAWI LDRAVIPASKIRGPRFLSVKRLMGLFMSFLAMASVMVESFRKKAISOQYAN NSNAVVDISAMMLVPOYVLGLAEALTAIGQTEFFTEPKSMSSAASLGLGMAVA SLASVLANAVNELTSRNGKESWVSDNINKHYNYIYVWLAINMSFINIYVICWSY GLPVDVRNGVNGVREELIDIVGKFEKEDLSPPVKTN" 26156. .26179 /rpt_family="(TA)n" 26779. .26899 /rpt_family="(CAAAA)n" join(<26899. .26982,27574. .>27987) /gene="F915.5" complement(<26899. .>27987 /gene="F915.5" /note="predicted by genscan" join(26899. .26982,27574. .27987) /gene="F915.5" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29405.1" /db_xref="GI:6850342" /translation="MMRYCQTPDYPSSNMDDLKAFEDETWYKGGKKEAWSLFDVLY SEEMKNMGEYTLINMYNECFKIGRFDQAIOFYKTKALKLNPDVTCYRNITTLNKO GVLFEEAYFEEMCSDRLVPPDVFTYTTMIDAYLKAGKTEDALRISKNMVDALFGQVA WLACL" complement(27450. .27491) /rpt_family="(TA)n" complement(join(<29019. .29311,29402. .29473,29582. .>29780)) /gene="F915.6" complement(<29019. .>29780) /gene="F915.6" /note="similar to PGP224 GB:AAD02554 from [Petunia x hybrida]" complement(join(29019. .29311,29402. .29473,29582. .29780))
repeat_region	complement(1497. .1591) /rpt_family="(TAAAA)n" complement(2125. .2194) /rpt_family="(GAAA)n" complement(join(<7284. .7610,7681. .8652,8869. .9016, 9063. .9373,9467. .9826,9904. .9947,10041. .10095, 10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" complement(<7284. .>10639) /gene="F915.1" /note="predicted by genefinder" complement(join(7284. .7610,7681. .8652,8869. .9016, 9063. .9373,9467. .9826,9904. .9947,10041. .10095, 10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29402.1" /db_xref="GI:6850339" /translation="MENNEATNGSKSSNSFVNKKRAEGFDITDKKRNLERKSKOL NPNTIAYAQIIGLQMDTSSVLLFFDKQRFINAGEGLQRFCTEHKIKLSKIDH VFLSRVCSSTAGLPGLLTLTLAGIGEGLSVNVGPGSDLYLVDMKSFIPRAAMVHT RSPGSSSTDPPIVLVNDVVKISAIILKPCSEDSGNKSGDLSVYVCPFLVLCPE DLEKAKKVFQPGPKYKSLQSGESVKSDESDITVHPSDVMGSPSLGPIVLLVDCPTE SHAELFSLKLSLEYSSPEDEQIGAFVNCIIHLSPPSVTSFTQSMKKFHLTQH ILAGHQRFLPILIVSHQKTVRKMAFPILKASSRAARLNLYLCPQFFPAPGFWPSOL TNSIITDPTSNKFNLPVAIRGIDRSCIPAPLTSSEVDELLEIPEIKDKSEIKQ FNKQHNKTIIEKLWLESECTVLPNCLEKIRDDMEIVILGTSSQPSKYRNYSALFI DLFSGSLDDCGEGTIGOLKKRYGLGDGADEAVRKLCIWSHIHADHTGLARILAL RSKLLAGVTHPEVIVGPRPLKFLDAYORLEDDMEFLDCRSTTASWASLESGEA EGSLFTQSGPMQSFYKRSIDMNSVYLCLKNLKLSEIGLNDLISFPVHCIPQAY GVVKAARVNSVGEQILGWKMDYSGDSRCPETVEASRDATILIHAEATFADALIEEA LAKNHTKEAIDVGAANRYVLTHFSQPKPIVIDESHMHTCIAFDLMSINNA DLHVLKPLVPYFKTLFRDEWVEDADDDVAMDLDKEAL" 10641. .10753 /rpt_family="(GAA)n" 10928. .11001 /gene="F915.2" /product="trna-Ala" /anticodon=(pos:10962. .10964,aa:Ala) 10928. .11001 /gene="F915.2" complement(join(11063. .11583,12194. .>12301)) /gene="F915.3" complement(11063. .>12301) /gene="F915.3" /note="similar to delta tonoplast integral protein GB:AAC49281 from [Arabidopsis thaliana]" complement(join(11317. .11583,12194. .>12301))	repeat_region	13177. .13298 /note="novel repeat from T25M19 22823 to 23610 788 nt"
repeat_region	complement(1497. .1591) /rpt_family="(TAAAA)n" complement(2125. .2194) /rpt_family="(GAAA)n" complement(join(<7284. .7610,7681. .8652,8869. .9016, 9063. .9373,9467. .9826,9904. .9947,10041. .10095, 10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" complement(<7284. .>10639) /gene="F915.1" /note="predicted by genefinder" complement(join(7284. .7610,7681. .8652,8869. .9016, 9063. .9373,9467. .9826,9904. .9947,10041. .10095, 10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29402.1" /db_xref="GI:6850339" /translation="MENNEATNGSKSSNSFVNKKRAEGFDITDKKRNLERKSKOL NPNTIAYAQIIGLQMDTSSVLLFFDKQRFINAGEGLQRFCTEHKIKLSKIDH VFLSRVCSSTAGLPGLLTLTLAGIGEGLSVNVGPGSDLYLVDMKSFIPRAAMVHT RSPGSSSTDPPIVLVNDVVKISAIILKPCSEDSGNKSGDLSVYVCPFLVLCPE DLEKAKKVFQPGPKYKSLQSGESVKSDESDITVHPSDVMGSPSLGPIVLLVDCPTE SHAELFSLKLSLEYSSPEDEQIGAFVNCIIHLSPPSVTSFTQSMKKFHLTQH ILAGHQRFLPILIVSHQKTVRKMAFPILKASSRAARLNLYLCPQFFPAPGFWPSOL TNSIITDPTSNKFNLPVAIRGIDRSCIPAPLTSSEVDELLEIPEIKDKSEIKQ FNKQHNKTIIEKLWLESECTVLPNCLEKIRDDMEIVILGTSSQPSKYRNYSALFI DLFSGSLDDCGEGTIGOLKKRYGLGDGADEAVRKLCIWSHIHADHTGLARILAL RSKLLAGVTHPEVIVGPRPLKFLDAYORLEDDMEFLDCRSTTASWASLESGEA EGSLFTQSGPMQSFYKRSIDMNSVYLCLKNLKLSEIGLNDLISFPVHCIPQAY GVVKAARVNSVGEQILGWKMDYSGDSRCPETVEASRDATILIHAEATFADALIEEA LAKNHTKEAIDVGAANRYVLTHFSQPKPIVIDESHMHTCIAFDLMSINNA DLHVLKPLVPYFKTLFRDEWVEDADDDVAMDLDKEAL" 10641. .10753 /rpt_family="(GAA)n" 10928. .11001 /gene="F915.2" /product="trna-Ala" /anticodon=(pos:10962. .10964,aa:Ala) 10928. .11001 /gene="F915.2" complement(join(11063. .11583,12194. .>12301)) /gene="F915.3" complement(11063. .>12301) /gene="F915.3" /note="similar to delta tonoplast integral protein GB:AAC49281 from [Arabidopsis thaliana]" complement(join(11317. .11583,12194. .>12301))	repeat_region	15264. .15339 /rpt_family="(GAAA)n" complement(18843. .18918) /rpt_family="(GAAA)n" 19668. .19760 /rpt_family="POLY_A" complement(20890. .20985) /rpt_family="(GA)n" complement(20931. .20985) /rpt_family="(GAA)n" join(<21044. .21140,21233. .21450,23180. .23739,24098. .25199) /gene="F915.4" complement(<21044. .25199 /gene="F915.4" /note="similar to PEPTIDE TRANSPORTER PTR2-B GB:P46032 from [Arabidopsis thaliana]" join(21044. .21140,21233. .21450,23180. .23739,24098. .25046) /gene="F915.4" /codon_start=1 /product="putative peptide transporter" /protein_id="AAF29404.1" /db_xref="GI:6850341" /translation="MENPNETEAKIQITNEGKTKGGIITMPFIIANEAFKVASVG LLPNMILYLRDYPFGVAKGTNVLPMASASNTPLLGAFLSDSYLGRLFTISLASL SFLGMVLLMTAMLPQVRKSPCDTPAAGSHCGSTASQALYSAFALISIGSGIRP CSLAFGADQDNKNERNVLEFSGYASSAVAVIAFTGIVYIOEHLGKIGFG VPALMLIAALLFLASPLYVTVGTGKSLFTGLAQATVAAYKKRKLSPDHPHSDFCY YHMKSEIKAPSKOKLRLNKAACLIISNREEIGSDGFPALNPWLCTTDKVELKALIKV IPIWSTGIMNSTSSQQLQATSMRRLSRHSGSFQVPGAGFMTIILALAWI LDRAVIPASKIRGPRFLSVKRLMGLFMSFLAMASVMVESFRKKAISOQYAN NSNAVVDISAMMLVPOYVLGLAEALTAIGQTEFFTEPKSMSSAASLGLGMAVA SLASVLANAVNELTSRNGKESWVSDNINKHYNYIYVWLAINMSFINIYVICWSY GLPVDVRNGVNGVREELIDIVGKFEKEDLSPPVKTN" 26156. .26179 /rpt_family="(TA)n" 26779. .26899 /rpt_family="(CAAAA)n" join(<26899. .26982,27574. .>27987) /gene="F915.5" complement(<26899. .>27987 /gene="F915.5" /note="predicted by genscan" join(26899. .26982,27574. .27987) /gene="F915.5" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29405.1" /db_xref="GI:6850342" /translation="MMRYCQTPDYPSSNMDDLKAFEDETWYKGGKKEAWSLFDVLY SEEMKNMGEYTLINMYNECFKIGRFDQAIOFYKTKALKLNPDVTCYRNITTLNKO GVLFEEAYFEEMCSDRLVPPDVFTYTTMIDAYLKAGKTEDALRISKNMVDALFGQVA WLACL" complement(27450. .27491) /rpt_family="(TA)n" complement(join(<29019. .29311,29402. .29473,29582. .>29780)) /gene="F915.6" complement(<29019. .>29780) /gene="F915.6" /note="similar to PGP224 GB:AAD02554 from [Petunia x hybrida]" complement(join(29019. .29311,29402. .29473,29582. .29780))
repeat_region	complement(1497. .1591) /rpt_family="(TAAAA)n" complement(2125. .2194) /rpt_family="(GAAA)n" complement(join(<7284. .7610,7681. .8652,8869. .9016, 9063. .9373,9467. .9826,9904. .9947,10041. .10095, 10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" complement(<7284. .>10639) /gene="F915.1" /note="predicted by genefinder" complement(join(7284. .7610,7681. .8652,8869. .9016, 9063. .9373,9467. .9826,9904. .9947,10041. .10095, 10177. .10221,10306. .10395,10478. .>10639)) /gene="F915.1" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29402.1" /db_xref="GI:6850339" /translation="MENNEATNGSKSSNSFVNKKRAEGFDITDKKRNLERKSKOL NPNTIAYAQIIGLQMDTSSVLLFFDKQRFINAGEGLQRFCTEHKIKLSKIDH VFLSRVCSSTAGLPGLLTLTLAGIGEGLSVNVGPGSDLYLVDMKSFIPRAAMVHT RSPGSSSTDPPIVLVNDVVKISAIILKPCSEDSGNKSGDLSVYVCPFLVLCPE DLEKAKKVFQPGPKYKSLQSGESVKSDESDITVHPSDVMGSPSLGPIVLLVDCPTE SHAELFSLKLSLEYSSPEDEQIGAFVNCIIHLSPPSVTSFTQSMKKFHLTQH ILAGHQRFLPILIVSHQKTVRKMAFPILKASSRAARLNLYLCPQFFPAPGFWPSOL TNSIITDPTSNKFNLPVAIRGIDRSCIPAPLTSSEVDELLEIPEIKDKSEIKQ FNKQHNKTIIEKLWLESECTVLPNCLEKIRDDMEIVILGTSSQPSKYRNYSALFI DLFSGSLDDCGEGTIGOLKKRYGLGDGADEAVRKLCIWSHIHADHTGLARILAL RSKLLAGVTHPEVIVGPRPLKFLDAYORLEDDMEFLDCRSTTASWASLESGEA EGSLFTQSGPMQSFYKRSIDMNSVYLCLKNLKLSEIGLNDLISFPVHCIPQAY GVVKAARVNSVGEQILGWKMDYSGDSRCPETVEASRDATILIHAEATFADALIEEA LAKNHTKEAIDVGAANRYVLTHFSQPKPIVIDESHMHTCIAFDLMSINNA DLHVLKPLVPYFKTLFRDEWVEDADDDVAMDLDKEAL" 10641. .10753 /rpt_family="(GAA)n" 10928. .11001 /gene="F915.2" /product="trna-Ala" /anticodon=(pos:10962. .10964,aa:Ala) 10928. .11001 /gene="F915.2" complement(join(11063. .11583,12194. .>12301)) /gene="F915.3" complement(11063. .>12301) /gene="F915.3" /note="similar to delta tonoplast integral protein GB:AAC49281 from [Arabidopsis thaliana]" complement(join(11317. .11583,12194. .>12301))	repeat_region	15264. .15339 /rpt_family="(GAAA)n" complement(18843. .18918) /rpt_family="(GAAA)n" 19668. .19760 /rpt_family="POLY_A" complement(20890. .20985) /rpt_family="(GA)n" complement(20931. .20985) /rpt_family="(GAA)n" join(<21044. .21140,21233. .21450,23180. .23739,24098. .25199) /gene="F915.4" complement(<21044. .25199 /gene="F915.4" /note="similar to PEPTIDE TRANSPORTER PTR2-B GB:P46032 from [Arabidopsis thaliana]" join(21044. .21140,21233. .21450,23180. .23739,24098. .25046) /gene="F915.4" /codon_start=1 /product="putative peptide transporter" /protein_id="AAF29404.1" /db_xref="GI:6850341" /translation="MENPNETEAKIQITNEGKTKGGIITMPFIIANEAFKVASVG LLPNMILYLRDYPFGVAKGTNVLPMASASNTPLLGAFLSDSYLGRLFTISLASL SFLGMVLLMTAMLPQVRKSPCDTPAAGSHCGSTASQALYSAFALISIGSGIRP CSLAFGADQDNKNERNVLEFSGYASSAVAVIAFTGIVYIOEHLGKIGFG VPALMLIAALLFLASPLYVTVGTGKSLFTGLAQATVAAYKKRKLSPDHPHSDFCY YHMKSEIKAPSKOKLRLNKAACLIISNREEIGSDGFPALNPWLCTTDKVELKALIKV IPIWSTGIMNSTSSQQLQATSMRRLSRHSGSFQVPGAGFMTIILALAWI LDRAVIPASKIRGPRFLSVKRLMGLFMSFLAMASVMVESFRKKAISOQYAN NSNAVVDISAMMLVPOYVLGLAEALTAIGQTEFFTEPKSMSSAASLGLGMAVA SLASVLANAVNELTSRNGKESWVSDNINKHYNYIYVWLAINMSFINIYVICWSY GLPVDVRNGVNGVREELIDIVGKFEKEDLSPPVKTN" 26156. .26179 /rpt_family="(TA)n" 26779. .26899 /rpt_family="(CAAAA)n" join(<26899. .26982,27574. .>27987) /gene="F915.5" complement(<26899. .>27987 /gene="F915.5" /note="predicted by genscan" join(26899. .26982,27574. .27987) /gene="F915.5" /note="hypothetical protein" /codon_start=1 /protein_id="AAF29405.1" /db_xref="GI:6850342" /translation="MMRYCQTPDYPSSNMDDLKAFEDETWYKGGKKEAWSLFDVLY SEEMKNMGEYTLINMYNECFKIGRFDQAIOFYKTKALKLNPDVTCYRNITTLNKO GVLFEEAYFEEMCSDRLVPPDVFTYTTMIDAYLKAGKTEDALRISKNMVDALFGQVA WLACL" complement(27450. .27491) /rpt_family="(TA)n" complement(join(<29019. .29311,29402. .29473,29582. .>29780)) /gene="F915.6" complement(<29019. .>29780) /gene="F915.6" /note="similar to PGP224 GB:AAD02554 from [Petunia x hybrida]" complement(join(29019. .29311,29402. .29473,29582. .29780))
repeat_region	complement(1497. .1591) /rpt_family="(TAAAA)n" complement(2125. .2194) /rpt_family="(GAAA)n" complement(join(<7284. .7610,761		

Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.

CDS are numbered using the following system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

FEATURES

location/Qualifiers

1..32329

/organism="Schizosaccharomyces pombe"

/strain="972h"

/db_xref="taxon:4896"

/chromosome="II"

/clone="cosmid c3D6"

/map="IIR"

1..359

/gene="SPBC3D5.01"

1..359

/partial

/gene="SPBC3D5.01"

/note="SPBC3D5.01, len:118aa"

/codon_start=3

/label="SPBC3D5.01"

/product="hypothetical protein"

/protein_id="CAB09128.1"

/db_xref="GI:2117297"

/db_xref="SWISS-PROT:P87166"

/translation="SFPYHYHFSPLVFLPHSFALPSTSISSRFLASPHQSLLFS
PAARHPSALHVLSPVRFAYHRCPRSVALTTPACRFVPPSSDVNHAIGRTSGREVF
NSLPSPVHGSDTEL"

1655..2827

/gene="SPBC3D6.02"

1655..2827

/gene="SPBC3D6.02"

/note="SPBC3D6.02, len:390, SIMILARITY:Schizosaccharomyces
pombe, 042663, hypothetical 39.7 kd protein., (383 aa),
fast a scores: opt: 301, E():9.8e-13, (22.7% identity in
396 aa)"

/codon_start=1

/label="SPBC3D6.02"

/product="hypothetical protein."

/protein_id="CAB09122.1"

/db_xref="GI:2117298"

/db_xref="SWISS-PROT:P87167"

/translation="MQLNLSFLGFAASTAVLASSADAAPTLYKRKSKSNFTAKSVAI
LDGVNNTFGVISTREKPPMDYHIWHTAPSTGKTYLPWNDVAPSTYV IDEGFLR
TGLKRAYLGDNMDLAFTNHTDKASHDDGTHRVLSORKNANFMATQRCGLDPFG
YOLKRSKVFQAGQVFGSGRSIDCWMDSVNLNRYGNTGALSPLPRNLISA
DIPAKSRLFPHGIVNENFAPNKRMRFTASEATTVNGQNSVLTVDVPRSLIY
TTSICALERFTNEFFPMVTPNDGSAQFVVMNSGNPKQTSTSSKGPRLYIVARENC
YTRGCEYTNPCPRAGHSHTLEAPASFNSTSIWQISTSPRIGTVLQVYNSANED"

complement(3885..5921)

/gene="SPBC3D6.03c"

complement(3885..5921)

/gene="SPBC3D6.03c"

/note="SPBC3D6.03c, len:678,

SIMILARITY:Schizosaccharomyces pombe, YATA_SCHPO,

hypothetical 90.6 kd protein cid4.10 in chromosome i.,

(809 aa), fasta scores: opt: 280, E():1.8e-10, (32.0%

identity in 494 aa)"

/codon_start=1

/label="SPBC3D6.03c"

/product="conserved hypothetical protein"

/protein_id="CAB09123.1"

/db_xref="GI:2117299"

/db_xref="SPTREMBL:P87168"

/translation="MKASLLVPRALLFGQLPPKYSWYSVKRWOSQLTFRNKSKRNT
NRIMLSVSSLNPDLSIAPLLCVDLDNRKYLIGSMGELTOMKFRSQASVYV
LMPPLSLOLNAGWITAGLFGYLOSSGIONTWGLHAPKPVISIIKKSHHLLFSGSPLRLD
LNSFSEONADATNSFYLDEPEFTIKGNIYSNWSFLSPNSKEAGVFNADKALALGV
PFGPNGLKAGEAVSLKSDGTTWYVHQVPPRKQYFYVLGCSLSALNQMSKHYD
SPDXVPTCIHILEKGIWGPYIKFLSHPKFSAQHFIISCIELASNNPFGNKGKN
VLACRDFAADIKPSTLDTOTLPENTVLAKEETSMVLYDECKLSESPSYSPVKLA
KFSSFNPLPFENEGYTLDLGTSATCPWRBSLSYSVAIDGTVMIDCGEGATSQF
FRQGTNTEPMLRKLKAIPIHLHSDHYLGLNLVQWNNKANTNNSMHINIIGPKFLW
QWLQRLKSPANLOALLNRIIFIATKETTLPQLTSDLSISSVTHINDSCLISH
TKYGLVSGDTRPNKELVKAGISLGLHSTFEDDLKHEAIOHQHSTASALSVAQ
SNKAKALILTHFSQSYDAFLPPDWTIYPKSKTIYANDGLOWQOQFOSKORETI"

complement(4239..4268)

/gene="SPBC3D6.03c"

/note="PS00152 ATP synthase alpha and beta subunits
signature"

complement(5628..5651)

/gene="SPBC3D6.03c"

/note="PS00017 ATP/GTP-binding site motif A (P-loop)"

join(6157..6159,6210..6450,6555..6577)

/gene="rps25-1"

/note="SPBC3D6.15"

join(6157..6159,6210..6450,6555..6577)

/gene="rps25-1"

/note="SPBC3D6.15, len:92"

/codon_start=1

/label="rps25"

/product="40s ribosomal protein s25"

/protein_id="CAB09129.2"

/db_xref="GI:4581499"

/translation="MAPKKKSKGKVKDKAQAHTVFDKSIIDRINKEVPAFKFISVS
LVDRMKINGSLARIAIRDLAERGVIQKVQDHSKQALYTRAAASA"

6160..6165

/note="gtaagt, splice donor sequence"

6192..6209

/note="ctaattccatattag, splice branch and acceptor
sequence"

6451..6554

/note="confirmed"

complement(join(6886..8804,8855..9005))

/gene="SPBC3D6.04c"

complement(join(6886..8804,8855..9005))

/gene="SPBC3D6.04c"

/note="SPBC3D6.04c, len:689, SIMILARITY:Saccharomyces
cerevisiae, MAD1_YEAST, spindle assembly checkpoint
component ma d1, (749 aa), fasta scores: opt: 454,
E():1.1e-16, (23.4% identity in 710 aa)"

/codon_start=1

/label="SPBC3D6.04c"

/product="hypothetical protein"

/protein_id="CAB09124.1"

/db_xref="GI:2117300"

/db_xref="SPTREMBL:P87169"

/translation="MSSKITYQATTSMADSPRDPFQSRSQLPRFLATSVKKPNLKKP
SVNSANETNPKLASLFOLENKNDLKRKELEFEQTELQRLKAEHEQKNSLQLR
LTLVKEQLEEQSTYQKLEEEVNEKEATQVKIHLLDAKWEIEALTKQIEKNDQAL
SKKNHEVSNQALOKNTNLNLEKFLADSREQLTKCKLEAAAEQQLSELVHHQQ
LEESTKVSSSIELEKINAKLQISELEKAAQERIEKLSSNNRNVEILKEKND
LESLKRYFEYRDVATLELENERIOTELNSWKSLITNELPTPEAVSNKLVPLQNTNA
NLGERSVSSLESQLSNKPANGLKAAHTELETKLKELHEQNRRQROKSLQATQ
EIDLRLNLSYDDREALISEKNTMKKLETEGLVKLVDEYKLKLESPVSLDVEDT
SDEVSLQRRKNEHKGADGYTELKKNHLLFOVKEKTNIEAFIREQILTESSLAT
LRQELAQTVEINSCVLRHSNPTIKYERIKAAQLEMLNAENSKALKEDEKVCCLP
IQSFIAERKALDKKAEVREKRIEFSVKSLEFRAEVLGFKYLDLPMFNGS
VAVTYSYREDNTAFIFDQESSTMKLVGNPSPGPEFLIRFWCDERKKTIPGMLAALT
ELLDKND"

complement(8805..8818)

/note="ctgacacctgtag, splice branch and acceptor
sequence"

complement(8849..8854)

misc_feature

misc_feature

gene

CDS

misc_feature

misc_feature

intron

gene

CDS

misc_feature

misc_feature

Sequence features of the regions of 4,504,864 bp covered by sixty p1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
20277480

JOURNAL MEDLINE REFERENCE

2 (bases 1 to 75289)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Direct Submission
Submitted (01-FEB-1999) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Department of
Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935,
Fax:81-438-52-3934)

JOURNAL TITLE

COMMENT

Address for correspondence: kaose@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MYA6
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
<http://compbio.ornl.gov/grail-1.3/>),
GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and
SplicePredictor (Volker Brendel, Stanford University,
<http://gremliu1.zool.iastate.edu/cgi-bin/sp.cgi>).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MSL1 and the 3' clone is MDC8.

FEATURES source

Location/Qualifiers
1..75289
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MYA6"
/clone_lib="Mitsui P1"
complement(join(4483..4904,4991..5106,5192..5636,
6082..6469))
/note="gene_id:MYA6.1
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01260.1"
/db_xref="GI:9279703"
/translation="MERGMNQCISLWHSQSPKPTPTMLDRALSRPHSDADLSASGE
STDSKTRPHYLLASNLSRIGHQWPCILALLFLVLLFLISVAFHSHSFVCIS
RDPAAIRGFFGLDSEDFGALGPWPCRSKHGKEVWTSKDLKGLFEFPIYETRP
IKNNYMGFDHISFGLWFMARLWPKDMMIESGAFKHSTWLRQAMPDTPMISLTPRH
PEKYLKGPAYVDGTYTFAGKDFDVGSDWKNVLRKHGIDLSRVIVFDDHONEL
KRLKALRAGFRHLIFEDNYDPTGTDHYSLRQICDOSHIRGGGSCFKSDRMRSK
RKRFWEKAVDTSELGCGETWGWGEMHDDFNHNTPILSYNQHFQNSRYVESILDVY
WELPVPAGSLTHQSRYPATPPIVADGKRHLRFQIRGLGRDKSDVFNQYTMYLE
ISKPGS"
complement(7397..8479)
/note="gb|AAAF25964.1
gene_id:MYA6.2
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01261.1"
/db_xref="GI:9279704"
/translation="MSKLPPEELALEILVRLSMKDLARPCVCKTWRDLNDPGTET
YDMSPAFVSVDKFNKIVPNPQVIGDHPVINKLDFPDOSMIDESVCVLDRLDGLCVT
LKNHTLMWNPSKOFKIVPNPQVIGDHPVINKLDFPDOSMIDESVCVLDRLDGLCVT
FFPRTGSGESIRLIPYDMYIDRRGTFLDQYLYIAYRSSADRFILCNLSTHEYRK
LPLPVNOCVTSWMLGVTQKLCITEYEMCKKEIRISVMEKTSWSKIISLSMSFSIS
VDRIYDQVEPVSFTRKNDLVVTFGTGNDHMEMPEERTKKKMFLYKTGNRSEVR
FCNPLAGFLCECVETLKVNRIFI"
join(10577..10657,10791..10941,11019..11131,11203..11292,

CDS

11366..11429,11555..11619,11726..11932)
/note="gene_id:MYA6.3
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01262.1"
/db_xref="GI:9279705"
/translation="MENLHMKWDRSSSHQTPLSFKGNKQIQGGQREVTFHFVSLPL
AIVPLTKNIEAFQNSVLGNDKDKPLKQSLAEKMSIFVTKPTFLTHTVVMKLE
NYESVVKQNIQVLVSFKSQIFSNVQALKNRPVIFRLRGECMSGLDKTRVLYA
PVEEVGHVLIIDAFENVFAGKAKSRLKHLATLMNAYSRYKSKKMDTFDAREIHKEF
ENKDMGTVLIAREHLSORYKYDPNGYFHCASLPPHK"
join(13499..13671,13836..13881,14227..14264,14340..14419,
14556..14623,14732..15011,15103..15215,15295..15360,
15467..15572,15667..15731,15855..15895,15984..16167)
/note="gene_id:MYA6.4
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01263.1"
/db_xref="GI:9279706"
/translation="MIFYQSNKTQNFIFELAMVGLKESDFVLFLFMHSHVVCVGLT
VYLGRVDVFNRRGSEKSDMGYKKQKMGASLIKIFIRGEGTTCOMKLEEMGVKII
LPSSRNKHISIEGSDVCVTKASKRIATIDENVRPSLDYSHFVSLPLAHPVLVD
KLVNQNSILGHISIASDKQDOQANRATTSVAVDLKANSETNOVNVVCIKIPISVPP
KAKSKSTLLDIGIEKSIKPIKSTFHLTVMLKLNKRNACDVKISIFPSVAKGL
DNKPVFIRLGLDCMRGLDTRVLYAPVEEIGDEGRLLRACRILDIITDAFVRKGL
VLEDAKQSLKHLVTVNARHRRKRNKKMETFDAREIHKQFGNEDWGEVLIQEAH
LSQRFVDQNGYRCGSIFFPPEQRA"
join(18398..18524,18628..18878,19278..19652)
/note="gb|AAC49281.1
gene_id:MYA6.5"
/codon_start=1
/evidence=not_experimental
/product="delta tonoplast intrinsic protein"
/protein_id="BAB01264.1"
/db_xref="GI:9279707"
/translation="MAGVAFGSDFSLSLASURAYLAEIFISTLLFVAGVGSIAIYAK
LTSDAALDPGLVAIVACHGFALFVAVAGTANISGHVNPVAVTFGLAVGQITVIGV
FYWIAQLGSGTAPLLKYVTGLVPTSHVAGLGSIEGVMEIITFALVYTVAT
AADPKGSLGTIAPLAIIGLVANTILAAQFSGGSMNPARSGPAVAGDFSGHWYWM
VGPLICGGLAGLILYGNVFMGSEHVPFLASADF"
complement(join(19955..20047,20150..20212,20303..20365,
20457..20513,20611..20721,20957..21184))
/note="gene_id:MYA6.6
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01265.1"
/db_xref="GI:9279708"
/translation="MGSVQLSGSLVASLPPNHSFHKTKLNKPSYFFRSKHNAAAT
KTVRAISTAPASOPPAADPEPPAVDFAFVHSLVLLDGTDPDVHRRANGOKLRDIM
LDSNTELYGPKSLNSCAGVCTCATCWEIVNGKELLNPTDIEKELKRKPKNRL
ACQTVGNPDSTGLVLIQOLPEWKAHWNIPNIPNDDLETST"
join(22261..22656,22735..22824,23038..23092,23187..23230,
23347..23742,23833..24098,24182..24326,24415..24465,
24561..25533,25635..25982)
/note="gb|AAF29402.1
gene_id:MYA6.7
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01266.1"
/db_xref="GI:9279709"
/translation="MLTSSMPQNLISLFGSPKSSSFFALLIPRPSLYPPFIASSPAP
SRPPTAGYRRSGSPPRKSSPEEQARKKSPMEKKAISFNSSSDFENKRRRA
ESGLDKVCKKLNKRNTRINTNTIAYVQILGTMDTQDTPSVLLFPDKORFIFNA
GVDHIFSRVCKTAGGLPGLLLTAGICEQLSVNMGPSDKLVDMKSRFIPRA
AMVHRYFQPSNLISDSAPQIGSLPKDDAYVLVDVVKISAILLEPRLSESSGPK
GTAVIYCELPENIGKDPKAMALGRAGPKYSVLSQSQSVKSDFKDITVHPSDVM
GSPGQPVLLVDDCTESHAEELLSPSKMITYYSCLDNSTDGNKLVNCLIHSPASVT
NSTYRSMKRFHSAQHILAGHAENKNEEPILRASSRITARNLCPQFPFAPGFWSH
QHDNNSINPITSLSKFDLSLGSISAENLLKFTLRPHGLNVDLRSSIPSRULTALRYMD

CDS

CDS

CDS

CDS

RESULT 15

AE001039 23533 bp DNA BCT 15-DEC-1997
LOCUS Archaeoglobus fulgidus section 68 of 172 of the complete genome.
DEFINITION AE001039 AE000782
ACCESSION AE001039.1 GI:2689362
KEYWORDS

SOURCE Archaeoglobus fulgidus.
ORGANISM Archaeoglobus fulgidus

REFERENCE 1 (bases 1 to 23533)
AUTHORS Klenk H.P., Clayton R.A., Tomb J., White, O., Nelson, K.E.,

Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
Richardson, D.L., Kierulff, A.R., Graham, D.E., Kyrpides, N.C.,
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.F., Glodek, A.,
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
Woese, C.R. and Venter, J.C.

TITLE The complete genome sequence of the hyperthermophilic,

sulphate-reducing archaeon Archaeoglobus fulgidus

Nature 390 (6558), 364-370 (1997)

98049343

2 (bases 1 to 23533)

Klenk, H.P., Clayton, R.A., Tomb, J., White, O., Nelson, K.E.,
Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D.,
Richardson, D.L., Kierulff, A.R., Graham, D.E., Kyrpides, N.C.,
Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton, G.G., Gill, S.,
Kirkness, E.F., Dougherty, B.A., McKenney, K., Adams, M.D., Loftus, B.,
Peterson, S., Reich, C.I., McNeil, L.K., Badger, J.F., Glodek, A.,
Zhou, L., Overbeek, R., Gocayne, J.D., Weidman, J.F., McDonald, L.,
Utterback, T., Cotton, M.D., Spriggs, T., Artach, P., Kaine, B.P.,
Sykes, S.M., Sadow, P.W., D'Andrea, K.P., Bowman, C., Fujii, C.,
Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith, H.O.,
Woese, C.R. and Venter, J.C.

Direct Submission

Submitted (15-DEC-1997) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

REMARK In order to show the genes in ascending order on the genome, the
origin of this version has been moved by TIGR to position 2093570
of the original version and the opposite strand is shown from the
original version.

COMMENT On Dec 16, 1997 this sequence version replaced gi:2649651.

FEATURES

source

1. 23533

/organism="Archaeoglobus fulgidus"

/db_xref="taxon:2234"

/complement(125..757)

/gene="AF0928"

/complement(125..757)

/gene="AF0928"

/note="similar to PID:1050822 PID:1164975 PID:1420338

percent identity: 27.60; identified by sequence

similarity; putative"

/codon_start=1

/transl_table=1

/product="conserved hypothetical protein"

/protein_id="AAB90310.1"

/db_xref="GI:2649668"

/translation="MHQVIOQRKPESSGESSRFFELAFDLDTLNIPIFDEIRREL
GIEGRFLESIMDEKDLQROEMLRVLESYELRCAMEAPYAYAVELVGLDAIKGVI
TNSRSEIIVAEIRLGFEDFLIGREDAPKPSPEPLNLRMFEDVSPSKALMWGDFL
FDLUSGKAAGAKTALITLDRNGMAKSFQIADYVFSKLADFLFEVQ"

/complement(699..1424)

/gene="AF0929"

/complement(699..1424)

/gene="AF0929"

/note="similar to GB:L77117 SP:Q57700 PID:1590982 percent

identity: 46.67; identified by sequence similarity;

CDS

gene

CDS

gene

CDS

gene

CDS

putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAB90309.1"
/db_xref="GI:2649667"
/translation="MYVONGIKLLRKALKINAGEKVIFTPVOOTLMKQLILALDVM
GEKAMEIAKVAEHVDRIKVNPLVLSAGVIMKRLSEIKPVIADFKIADVPYTSLLI
ARIAPENAESVIVHGFYSDTLREVCRVAEEFGKVAVTLSLSPGGEFMSAVSLK
IVERAKAEGCHVLIAPSTRIERLIRKAAAGDMETLCPGIAQKGSIEAVKYADGIIV
GRGIYASGNPAEARKLRVLKI"
1430..2674
/gene="AF0930"
1430..2674
/gene="AF0930"
/note="similar to GB:L77117 SP:Q58080 PID:1591380 percent
identity: 44.81; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="molybdenum cofactor biosynthesis protein
(moaA-2)"
/protein_id="AAB90308.1"
/db_xref="GI:2649666"
/translation="MYREKFWKRRVDEALYRLKKKALETEDSVSVNLQGRVLAE
DIVAGDFVPEKFNRAAMDGYAKAEDTFGASVSNPIMQLAGSVEIGEYKVVESGMA
VRITGAMMPECTNAVMLEHTRLNGNEVEVLKSVPMKNSRVSGDEAAGAEVFRKG
ETLQPDAGVIAISLGIESRVVRKPRVAITVGNELAEVGELEGAKIYNSNPMICN
ALLERGFPAISIGIARDDEELRNKLOKALEPDAVITGTSVGAADLVDPVVGEGE
IVFHVSGMKPMGTAAGVEGKPEMLPGSPAPALLGFTYFAIPALYRMVNRITARK
WSRGVQLQGRIPSEIGVRSNVRVLWEDGKVPIRISGSLISFVRANALLVVPEDK
EGYEGEVEVTLRLDRDTEVFE"
2671..4635
/gene="AF0931"
2671..4635
/gene="AF0931"
/note="similar to GB:L77117 SP:Q58296 PID:1591564 percent
identity: 50.83; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="molybdenum cofactor biosynthesis protein
(moaA-1)"
/protein_id="AAB90307.1"
/db_xref="GI:2649665"
/translation="MRKVFRDVTVDVDEARGLLFNHFSPRKTVEVPLHLAAGRVLAKD
VYALTDVPFORATMDGYAVRAEDTFEAEEDSPVTLKVGTVEACKRPELEVESKAV
EIATGMMPKGANAVVMVVEFTRRREDEIDIFKVPSPENVTAGSDVWAGELIAREGT
ELTHREIGVMAACGISEVEYCRPKVAIISTGNELVKPGERLGEKIDVNSYIAAA
VEENGGEAILMGIAKDNEEMRSIIKALREADIVITSGTSAGADVMYRILDDFHP
GVIVHGIATKPKPAVIAVSDSKPIFGLPYPTSAMTVFEVIVAPIIRELAGRESSG
AKVRARLAIKVFSAEGRREYLPVNVVEGAGSYAVPSGYSYCAVTAFTDGFVEIP
ENVVMLEGEDEVEVKLFSLKPADLMIIGCHIGIDVILLMRKKPKPYTSKVVNGVST
GGILAVKRGEDIAGTHLLDSEGVYNEPIRVRYGVDRADVLVKGYREGLIVAKGNPK
SITGFEDLLREDVTFINNRNSGTRILDMYLRVAREKRGSLNFELTASIKGYSLEAK
THTSVAVASGADGVGIRSVASQYGLDFILRSEYDFILIRDRLEKKAVIDFLE
CLTSEFAKELERVVGLRVYERTREVIED"
4677..6677
/gene="AF0932"
4677..6677
/gene="AF0932"
/note="similar to GB:L77117 SP:Q58010 PID:1591298 percent
identity: 34.43; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAB90306.1"
/db_xref="GI:2649664"
/translation="MRLLEHESKALLEKYGITAKCIFCETEEAOVKAKEIGPPV
MKVAGREILHKSDCCVILNVKSEDEVEVFORLMSIFKAEGVNLQPOLEKGIIEVIG
VAENQFGSVAMFGLGGVFEVVKDVSFLLPTRDAEEMVVRVKKLLEGGYGVK
GDVGAVDILLKLENEVEKEGIVEMDLNPFVYFVYRGAVVVDARIYVGERKRFEMDVG

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2001, 05:20:16 ; Search time 128.59 Seconds
(without alignments)
7247.991 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atcggcgcttgcgcgt.....agaaggtcagagccagtgta 2481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 480022 seqs, 18781343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_36: *
1: /cgnl_8/gcgdata/geneseq/geneseqn/NA1980.DAT: *
2: /cgnl_8/gcgdata/geneseq/geneseqn/NA1981.DAT: *
3: /cgnl_8/gcgdata/geneseq/geneseqn/NA1982.DAT: *
4: /cgnl_8/gcgdata/geneseq/geneseqn/NA1983.DAT: *
5: /cgnl_8/gcgdata/geneseq/geneseqn/NA1984.DAT: *
6: /cgnl_8/gcgdata/geneseq/geneseqn/NA1985.DAT: *
7: /cgnl_8/gcgdata/geneseq/geneseqn/NA1986.DAT: *
8: /cgnl_8/gcgdata/geneseq/geneseqn/NA1987.DAT: *
9: /cgnl_8/gcgdata/geneseq/geneseqn/NA1988.DAT: *
10: /cgnl_8/gcgdata/geneseq/geneseqn/NA1989.DAT: *
11: /cgnl_8/gcgdata/geneseq/geneseqn/NA1990.DAT: *
12: /cgnl_8/gcgdata/geneseq/geneseqn/NA1991.DAT: *
13: /cgnl_8/gcgdata/geneseq/geneseqn/NA1992.DAT: *
14: /cgnl_8/gcgdata/geneseq/geneseqn/NA1993.DAT: *
15: /cgnl_8/gcgdata/geneseq/geneseqn/NA1994.DAT: *
16: /cgnl_8/gcgdata/geneseq/geneseqn/NA1995.DAT: *
17: /cgnl_8/gcgdata/geneseq/geneseqn/NA1996.DAT: *
18: /cgnl_8/gcgdata/geneseq/geneseqn/NA1997.DAT: *
19: /cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT: *
20: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT: *
21: /cgnl_8/gcgdata/geneseq/geneseqn/NA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	237	9.6	238	21	280231 Human colon cancer
2	51.4	2.1	1501	21	AI2583 cDNA encoding a me
3	48.2	1.9	799	19	V55831 Nucleotide sequenc
4	48.2	1.9	1925	20	X90924 Epstein Barr Virus
5	48.2	1.9	5452	20	X90923 Anti-sense strand
6	48.2	1.9	8705	20	Z23778 Vector pShuttle DN
7	48.2	1.9	9600	19	V21683 Vector plasmid pCM
8	48.2	1.9	10380	20	Z22248 Nucleotide sequenc
9	48.2	1.9	10596	14	Q51731 Plasmid pCISBON f
10	48.2	1.9	10596	17	T40348 Plasmid pCISBON f
11	48.2	1.9	10596	20	X15650 Nucleotide sequenc
12	45	1.8	795	19	V55830 FLGA insert stabil

13	43.8	1.8	1528	20	X35555 Secreted protein c
14	43.6	1.8	217	16	Q97711 Branched-chain alp
15	43.4	1.7	1567	19	X03145 Aspergillus terreu
16	43	1.7	2125	20	X36279 Wheat Rht clone 5a
17	42.8	1.7	453	20	X36263 DNA sequence obtai
18	42.8	1.7	2709	20	X36275 Composite DNA sequ
19	42.8	1.7	16442	18	X83006 Partial mouse WRN
20	42.6	1.7	10732	21	AI0594 Gene encoding a su
21	42.4	1.7	324	20	X36273 DNA sequence obtai
22	42	1.7	2297	21	X93309 Sequence encoding
23	41.4	1.7	38186	20	Z32028 Human MTH1 relate
24	41	1.7	2301	19	V20445 Human trk oncogene
25	41	1.7	2301	21	A09303 Human trk oncogene
26	40.8	1.6	1429	21	Z93300 Sequence encoding
27	40.4	1.6	15377	13	Q25975 MH mutant porcine
28	40.2	1.6	1412	17	T26989 Thermus aquaticus
29	39.8	1.6	6306	15	Q54841 Sequence encoding
30	39.8	1.6	6579	18	T77783 cDNA encoding LexA
31	39.8	1.6	6819	18	T77782 cDNA encoding GAL4
32	39.6	1.6	730	19	V62161 HSV-2 strain SB5 C
33	39.6	1.6	925	21	Z52534 Human secreted pro
34	39.6	1.6	2744	16	Q98470 Misp1-containing p
35	39.6	1.6	4289	19	V62147 HSV-2 strain SB5 C
36	39.6	1.6	9551	20	Z22301 cDNA encoding a hu
37	39.6	1.6	51259	18	X83007 Partial mouse WRN
38	39.6	1.6	117213	19	V62176 HSV-2 strain SB5 C
39	39.4	1.6	1675	20	Z32195 Corn hexose carrie
40	39.4	1.6	4020	18	T91361 Orf virus genomic
41	39.4	1.6	114955	20	X53491 Human adenosine A1
42	39	1.6	1994	21	A35027 Human adenosine re
43	39	1.6	2009	20	Z07513 Human RIP-associat
44	39	1.6	2034	20	Z07514 Human RIP-associat
45	39	1.6	8631	21	A35028 Human adenosine re

ALIGNMENTS

RESULT	1
280231	
ID	Z80231 standard; cDNA; 238 BP.
XX	Z80231;
AC	Z80231;
DT	07-APR-2000 (first entry)
XX	Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
DE	Human; gene expression product; diagnosis; tumour; colon cancer;
KW	colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW	cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW	hyperplasia; ds.
XX	Homo sapiens.
OS	XX
PN	WO9964576-A2.
XX	XX
PD	16-DEC-1999.
XX	XX
PF	09-JUN-1999; 99WO-IB01062.
XX	XX
PR	10-JUN-1998; 98US-0088801.
XX	XX
PA	(FARB) BAYER CORP.
XX	XX
PI	Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI	Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JR;
PI	Schlegel R;
XX	WPI; 2000-087220/07.
DR	XX
XX	Novel nucleic acids, used to develop products for the diagnosis and
PT	treatment of disorders involving unwanted cell proliferation,

OS	Epstein-barr virus.
XX	
PN	WO9822577-Al.
XX	
PD	28-MAY-1998.
XX	
PF	17-NOV-1997; 97WO-IB01508.
XX	
PR	25-JUN-1997; 97US-0048945.
PR	15-NOV-1996; 96US-0030986.
XX	
PA	(MASU/) MASUCCI M G.
XX	
PI	Masucci MG;
XX	
DR	WPI; 1998-312463/27.
XX	
PT	New fusion proteins resistant to proteolytic degradation -
PT	comprising a core protein with a stabilising polypeptide comprising
PT	a peptide sequence containing glycine repeats
XX	
PS	Disclosure: Fig 4B; 120pp; English.
XX	
CC	This is a nucleotide sequence of the stabilising sequence-encoding
CC	insert. The invention provides a method for increasing the resistance
CC	of a core protein to proteolytic degradation that comprises linking or
CC	inserting onto or into the core protein a stabilising polypeptide of
CC	formula (Glya)(Glyb)(Glyc)2n where Glya, Glyb, Glyc are 1-6
CC	sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,
CC	Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not
CC	be identical from n repeat to n repeat. Alternatively a nucleic acid
CC	encoding the stabilising polypeptide can be linked onto or inserted into
CC	a nucleic acid encoding a core protein. The fusion proteins of the
CC	invention are more resistant to degradation by proteases and, thus, have
CC	a longer half-life than the unfused core protein. The products can be
CC	used for treating autoimmune diseases, cancer and inflammation. In
CC	particular, the core protein may be an IkappaB regulator protein for the
CC	treatment of inflammatory bowel disease, or a nitroreductase protein
CC	which can activate nitro drugs in enzyme/prodrug therapy to treat cancer
CC	or other pathological conditions. The fusion proteins can also be used in
CC	diagnostic methods such as in vivo imaging.
XX	
SQ	Sequence 799 BP; 201 A; 106 C; 479 G; 13 T; 0 other;

KW episome; transfection; origin of replication; EBV orip; receptor;
 KW eucaryotic host cell; recombinant cell line; ion channel; gene therapy;
 KW multiple gene expression; transporter protein; transcription factor;
 KW adhesion molecule; antisense therapy; gene amplification;
 KW cell immortalisation; ds.
 XX
 XX Epstein-barr virus.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1925
 FT FT /*tag= a
 FT FT /product= "EBNA 1"
 FT FT /transl_except= (pos:799..800, aa:Gly)
 FT FT /note= "The sequence is described throughout the
 FT FT specification as being 1926 nucleotides long, but a
 FT FT sequence of only 1925 bp has been given in figure 2"
 XX
 XX WO9947647-A1.
 PN
 XX
 XX
 PD 23-SEP-1999.
 XX
 XX 12-FEB-1999; 99WO-US03307.
 PF
 PF 18-MAR-1998; 98US-0040961.
 PR 06-AUG-1998; 98US-0130114.
 XX
 XX (PHAR-) PHARMACOPETA INC.
 PA
 XX
 PI Damaj BB, Horlick RA, Robbins AK;
 XX
 DR WPI; 1999-610610/52.
 DR P-PSDB; Y28843.
 XX
 XX New method for expressing genes from recombinant eukaryotic cells,
 PT useful for gene therapy -
 PT
 XX
 XX Claim 24; Fig 2; 86pp; English.
 PS
 PS
 XX
 CC The present sequence is a DNA encoding Epstein Barr Virus Nuclear
 CC Antigen 1 (EBNA 1), which is obtained from commercially available
 CC plasmid pCMVEBNA. EBNA 1 protein is used to stably maintain episomes
 CC containing EBV origin of replication (orip) and a gene encoding
 CC protein or RNA of interest. Eucaryotic host cells expressing EBNA 1
 CC protein are transfected with these episomes to produce recombinant
 CC cell lines expressing multiple genes of interest. This provides a
 CC rapid and reliable method of stably expressing multiple genes in
 CC transfected cells. The episomes are useful in the transfection of genes
 CC encoding receptors, transporter proteins, ion channels, adhesion
 CC molecules and transcription factors. The episomes carrying desired genes
 CC can also be used to transfect cells in gene therapy, antisense therapy,
 CC for gene amplification, cell immortalisation, etc.
 XX
 XX Sequence 1925 BP; 486 A; 352 C; 872 G; 215 T; 0 other;

nucleic acid (see also V21684-86) into a cell. The polycationic agent can condense with the nucleic acid and inhibit serum and/or nuclease degradation of the nucleic acid. The nucleic acid can be a vector, may express a therapeutic protein or a vaccinating viral or cancer antigen, or is itself therapeutic (antisense or ribozyme). The methods and compositions can be used in the gene therapy of many diseases.

Sequence 9600 BP; 2326 A; 2376 C; 2817 G; 2081 T; 0 other;

Query Match 1.9%; Score 48.2; DB 19; Length 9600;
Best Local Similarity 58.9%; Pred. No. 0.021;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps

QY 2327 aggaatggaggagccgaggaagaagcggcgagctgcggcaggtcgcgggccctctgt 2386
||||| ||||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 873 aggaacgaggaaggcgaggaaggcgagcgagcgagggcgagggcgagggcgaggg 932
||||| ||||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY 2387 ccaggggactggcagcgcgccctggagatggggagcctcagcagaagcgggccacacag 2446
||||| ||||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 933 ggcaggagcaggaaggcgaggaagcaggaaggcgagggcgagggcgagggcaggagg 992
||||| ||||| | |||| | |||| | |||| | |||| | |||| | |||| |
QY 2447 aggagccacagggccaagaagg 2467
||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |
Db 993 agggcgaggaaggcgaggaagg 1013
||||| |||| | |||| | |||| | |||| | |||| | |||| | |||| |

RESULT 8
222248
ID 222248 standard; DNA; 10380 BP.
AC 222248;
XX
DT 29-NOV-1999 (first entry)
DE Nucleotide sequence of pCEP4 vector.
KW
KW G_i protein coupled receptor; G_iPCR; G_i_a protein; cytoplasmic calcium;
KW calcium mobilization; nociceptin receptor; chemokine receptor; CCR3;
KW CCR2; interleukin 8 receptor type B; episomal expression vector; ss.
XX
OS Synthetic.
OS
PN WO9947921-A1.
XX
PD 23-SEP-1999.
XX
PF 09-FEB-1999; 99WO-USO2852.
XX
PR 18-MAR-1998; 98US-0040961.
XX
PA (PHAR-) PHARMACOPEIA INC.
XX
PI Damaj BB, Horlick RA, Robbins AK;
XX
DR WPI; 1999-562218/47.
XX
PT A method to enhance signal transduction responses involving cytoplasmic
XX calcium release .
XX
PS Examples; Fig 3A-J; 64pp; English.
XX
CC The invention relates to a cell transfected with a gene encoding a G_i
CC protein coupled receptor (G_iPCR) and with a gene coding for a G_i_a
CC protein capable of coupling to the G_iPCR, to increase the cytoplasmic
CC calcium upon binding of an agonist to the G_iPCR. The cell is used to
CC assay a test compound for its effect on a G_iPCR. The method and cells
CC are useful for assaying agonist and antagonist compounds, which bind to
CC G_iPCR which, stimulate intracellular calcium mobilization, such as the
CC nociceptin receptor, and particularly chemokine receptors such as CCR3,
CC CCR2 and the interleukin 8 receptor type B. The present sequence
CC represents the nucleotide sequence of the pCEP4 vector

```

XX SQ Sequence 10380 BP; 2560 A; 2561 C; 2964 G; 2295 T; 0 other;

Query Match          1.9%; Score 48.2; DB 20; Length 10380;
Best Local Similarity 58.9%; Pred. No. 0.022;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2327 agagatggagggcagggaggaagcggagctgcgcaggtgcggcgccctctgt 2386
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6359 aggagcaggaggaggcaggaggggcaggagcaggagggcaggaggcaggagg 6418

QY 2387 ccaggagctggcaggcgccctggagatggggagcctcagcaagcgggccacacag 2446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6419 ggcaggagcaggaggaggcaggagcaggaggaggggcaggaggcaggaggagg 6478

QY 2447 aggagccacagggcagaagg 2467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6479 aggggacaggaggcaggagg 6499

RESULT 9
Q51731
ID Q51731 standard; DNA; 10596 BP.
XX AC Q51731;
XX
XX 31-MAY-1994 (first entry)
XX
XX Plasmid pCisEBON for subcloning huHGF variants.
XX
XX Hepatocyte Growth Factor; HGF; variant; mutein; in vitro mutagenesis;
KW proteolysis resistant; liver; malignancy; CMV-driven;
KW Cytomegalovirus; episomal expression plasmid; ss.
XX
XX Synthetic.
XX
XX Key
FH enhancer Location/Qualifiers
FT 1..611
FT /*tag= a
FT /*note= "CMV enhancer/promoter"
FT 758..775
FT /*tag= b
FT /*label= SP6_promoter
FT 845..849
FT /*tag= c
FT /*note= "SP6 RNA start"
FT 902..966
FT /*tag= d
FT /*function= cloning_linker
FT 967..1107
FT /*tag= e
FT /*note= "SV40 poly A"
FT 1108..1531
FT /*tag= f
FT /*function= SV40_origin
FT 1580..4189
FT /*tag= g
FT /*label= EBNA-1
FT 4190..6374
FT /*tag= h
FT /*function= orip
FT 4295..4887
FT /*tag= i
FT /*note= "family of repeats"
FT 5866..5978
FT /*tag= j
FT /*note= "dyad region"
FT 6375..6457
FT /*tag= k
FT /*label= HSV_TK_terminator_3'-end
FT 6975..7975
FT /*tag= l

```

```

FT FT /phenotype= neomycin_resistance
FT FT /note= "tn5 neomycin phosphotransferase gene"
FT FT 7975..8112
FT FT /*tag= m
FT FT /*label= TK_promoter
FT FT 8114..8594
FT FT /*tag= n
FT FT /*function= M13_ori
FT FT 8595..10414
FT FT /*tag= o
FT FT /*label= delta_2a
XX
XX W09323541-A.
XX
XX 25-NOV-1993.
XX
XX 17-MAY-1993; 93WO-US04648.
XX
XX 18-MAY-1992; 92US-0884811.
XX 18-MAY-1992; 92US-0885971.
XX (GETH ) GENENTECH INC.
XX
XX Godowski PJ, Lokker NA, Mark MR;
XX WPI; 1993-386573/48.
XX
XX Hepatocyte growth factor variants - are resistant to proteolytic
PT cleavage into its two-chain form, used to treat malignancies
PT associated with HGF receptor
XX
XX Example 1; Fig 6; 87pp; English.
XX
XX Plasmid pCisEBON (a pRK5 derivative) is an episomal CMV driven
CC expression plasmid. HuHGF variants with enhanced receptor binding
CC activity were produced by site-directed mutagenesis. Stable
CC populations of preferred HGF variants were obtained by transfecting
CC human embryonic kidney 293 cells and then these were subcloned in
CC pCisEBON. See R52940-R52949 for examples of pref. HGF variants.
XX
XX Sequence 10596 BP; 2625 A; 2571 C; 3024 G; 2376 T; 0 other;

Query Match          1.9%; Score 48.2; DB 14; Length 10596;
Best Local Similarity 58.9%; Pred. No. 0.022;
Matches 83; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 2327 aggagatggagggcagggaggaagcggagctgcgcaggtgcggcgccctctgt 2386
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2407 aggagcaggaggaggaggcaggaggggcaggagcaggaggaggcaggaggagg 2466

QY 2387 ccaggagctggcaggcgccctggagatggggagcctcagcaagcgggccacacag 2446
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2467 ggcaggagcaggaggaggaggcaggagcaggaggaggggcaggaggcaggaggagg 2526

QY 2447 aggagccacagggcagaagg 2467
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2527 aggggacaggaggcaggagg 2547

RESULT 10
T40348
ID T40348 standard; DNA; 10596 BP.
XX AC T40348;
XX
XX 09-DEC-1996 (first entry)
XX
XX Plasmid pCisEBON for expression of hepatocyte growth factor.
XX
XX Human; hepatocyte growth factor; HGF; huHGF; serum; proteolytic cleavage;
KW pro-hormone; beta subunit; alpha subunit; kringle domain; prothrombin;
KW plasminogen; catalytic domain; serine protease; HGF variant;

```


[illegible]

Db 827 AGGAGCAGGAGGCGGCAGGAGCGGGGCAGGAGCGGGGCAGGAGG 886

Query Match 1.9%; Score 48.2; DB 1; Length 10596;

Query Match 1.9%; Score 48.2; DB 1; Length 10596;

	Query Match	1.98;	Score 48.2;	DB 1;	Length 10596;
	Best Local Similarity	58.9%;	Pred. No. 0.0043;		
	Matches 83;	Conservative 0;	Mismatches 58;	Indels 0;	Gaps 0;
QY	2327	aggagatggaggagcgcagggagaagcggagagctgcgcaggtgcgggcgcctcctgt	2386		
Db	2407	AGGAGCAGGAGGAGGGGCAGGAGGGCGAGGAGGAGGGCAGAGGGGCGAGGAGG	2466		
QY	2387	ccaggagagctggcaggcgccctggagatggggagcctcagcagaagcgggccccacacag	2446		
Db	2467	GGCAGGAGCAGGAGGAGGGCCAGGAGCAGGAGGAGGAGGGCGAGGAGGAGGAGG	2526		
QY	2447	aggagccacagggccaaagaag	2467		

	Query Match	1.98;	Score 48.2;	DB 1;	Length 10596;
	Best Local Similarity	58.9%;	Pred. No. 0.0043;		
	Matches 83;	Conservative 0;	Mismatches 58;	Indels 0;	Gaps 0;
QY	2327	aggagatggaggagcgcagggagaagcggagagctgcgcaggtgcgggcgcctcctgt	2386		
Db	2407	AGGAGCAGGAGGAGGGGCAGGAGGGCGAGGAGGAGGGCAGAGGGGCGAGGAGG	2466		
QY	2387	ccaggagagctggcaggcgccctggagatggggagcctcagcagaagcgggccccacacag	2446		
Db	2467	GGCAGGAGCAGGAGGAGGGCCAGGAGCAGGAGGAGGAGGGCGAGGAGGAGGAGG	2526		
QY	2447	aggagccacagggccaaagaag	2467		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 18, 2001, 01:12:05 ; Search time 1827.78 Seconds
(without alignments)
9511.854 Million cell updates/sec

Title: US-09-434-382-1
Perfect score: 2481
Sequence: 1 atgtggcgcttgcgcgt.....agaaggtcagagccacgtga 2481

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estov1:*
68: em_estov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_est41:*
88: gb_est42:*
89: gb_est43:*
90: gb_est44:*
91: gb_est45:*
92: gb_est46:*
93: gb_est47:*
94: gb_est48:*
95: gb_est49:*
96: gb_est50:*
97: gb_est51:*
98: gb_est52:*
99: gb_est53:*
100: gb_est54:*
101: gb_est55:*
102: gb_est56:*
103: gb_est57:*
104: gb_est67:*
105: gb_est68:*
106: gb_est69:*
107: gb_est70:*
108: gb_est71:*
109: gb_est72:*
110: gb_est73:*
111: gb_est74:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

117: em_estpl6:*
118: em_estpl7:*
119: em_estpl8:*
120: em_estro14:*
121: em_estro15:*
122: em_estro16:*
123: em_estro17:*
124: em_estro18:*
125: em_estro19:*
126: gb_est58:*
127: gb_est59:*
128: gb_est60:*
129: gb_est61:*
130: gb_est62:*
131: gb_est63:*
132: gb_est64:*
133: gb_est65:*
134: gb_est66:*
135: gb_est67:*
136: gb_est76:*
137: gb_est77:*
138: gb_est78:*
139: gb_est79:*
140: gb_est80:*
141: gb_est81:*
142: gb_est82:*
143: gb_est83:*
144: gb_est84:*
145: gb_est85:*
146: gb_est86:*
147: gb_est87:*
148: gb_gss1:*
149: gb_gss2:*
150: gb_gss3:*
151: gb_gss4:*
152: em_gss1:*
153: em_gss2:*
154: em_gss3:*
155: em_gss4:*
156: gb_gss5:*
157: gb_gss6:*
158: em_gss5:*
159: em_gss6:*
160: gb_gss7:*
161: gb_gss8:*
162: gb_gss9:*
163: em_gss7:*
164: em_gss8:*
165: gb_gss11:*
166: gb_gss10:*
167: em_gss9:*
168: em_gss10:*
169: em_gss11:*
170: em_gss12:*
171: gb_gss12:*
172: gb_gss13:*
173: gb_gss14:*
174: gb_gss15:*
175: gb_gss16:*
176: gb_gss17:*
177: gb_gss18:*
178: gb_gss19:*
179: em_gss13:*
180: gb_gss20:*
181: gb_gss21:*
182: gb_gss22:*
183: gb_gss23:*
184: gb_gss24:*
185: em_gss14:*
186: em_gss15:*
187: em_gss16:*
188: em_gss17:*
189: em_gss18:*

190: gb_gss25:*
191: gb_gss26:*
192: gb_gss27:*
193: gb_gss28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	709.6	28.6	761	135	BE795820	BE795820 601590856
2	679	27.4	823	110	BE619259	BE619259 601473130
3	676.6	27.3	984	135	BE744876	BE744876 601576324
4	676	27.2	676	106	BE260495	BE260495 601150702
5	662.6	26.7	812	136	BE867512	BE867512 601443010
6	655.6	26.4	735	137	BE902696	BE902696 601677393
7	651.4	26.3	688	135	BE794311	BE794311 601571442
8	649	26.2	949	135	BE744197	BE744197 601571168
9	641.6	25.9	992	135	BE747163	BE747163 601577254
10	625.6	25.2	745	137	BE900936	BE900936 601674206
11	622.8	25.1	664	107	BE383336	BE383336 601298249
12	610.8	24.6	692	107	BE382353	BE382353 601298656
13	610.4	24.6	790	137	BE892893	BE892893 601435738
14	592	23.9	677	135	BE742908	BE742908 601574609
15	578	23.3	940	135	BE743831	BE743831 601577742
16	576.4	23.2	612	106	BE304720	BE304720 601106236
17	553.8	22.3	713	109	BE537086	BE537086 601063325
18	543.4	21.9	938	106	BE260626	BE260626 601146116
19	516	20.8	531	4	AA243700	AA243700 zr68g08.s
20	501.2	20.2	574	105	BE250309	BE250309 600943455
21	495.2	20.0	511	95	AW771657	AW771657 hm59h06.x
22	492	19.8	493	92	AW575677	AW575677 UI-HF-BM0
23	480	19.3	512	87	AW175923	AW175923 RCI-BR007
24	478.6	19.3	691	107	BE409312	BE409312 601300940
25	470.4	19.0	536	87	AW206103	AW206103 UI-H-BI1-
26	468.4	18.9	493	7	AA455121	AA455121 zx78c04.s
27	463	18.7	474	4	AA291670	AA291670 zt37d04.s
28	460.8	18.6	588	3	AA184645	AA184645 mt58d10.r
29	460	18.5	501	5	AA311855	AA311855 EST182568
30	451	18.2	451	17	AI200296	AI200296 qf66b12.x
31	447.2	18.0	872	110	BE615669	BE615669 601279347
32	441	17.8	855	146	W27286	W27286 28h1 Human
33	436.6	17.6	617	137	BE914535	BE914535 601655633
34	435.4	17.5	446	10	AA634909	AA634909 ab27h02.r
35	430	17.3	431	5	AA310236	AA310236 EST181085
36	429	17.3	452	10	AA679618	AA679618 ag72c12.s
37	427.2	17.2	975	105	BE250412	BE250412 600943455
38	421	17.0	421	20	AI468143	AI468143 tf92g05.x
39	404	16.3	404	88	AW304130	AW304130 xs13e05.x
40	403.6	16.3	518	144	R87541	R87541 ym89b04.r1
41	402.4	16.2	416	106	BE298273	BE298273 601118143
42	397.8	16.0	627	107	BE386924	BE386924 601274815
43	390.4	15.7	698	135	BE795434	BE795434 601592991
44	389.8	15.7	457	14	AA994126	AA994126 ou38b06.s
45	386	15.6	422	146	W37591	W37591 zc10f03.r1

ALIGNMENTS

RESULT 1
BE795820 BE795820 761 bp mRNA EST 20-SEP-2000
LOCUS 601590856f1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
DEFINITION mRNA sequence.
ACCESSION BE795820
VERSION BE795820.1 GI:10217018
KEYWORDS EST.
SOURCE human.

```
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM802 row: p column: 14
High quality sequence stop: 761.

FEATURES
Source
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3945085"
/clone.lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_lines="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCACAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 172 a 206 c 219 g 164 t

Query Match 28.6%; Score 709.6; DB 135; Length 761;
Best Local Similarity 98.3%; Pred. No. 4.4e-175;
Matches 749; Conservative 0; Mismatches 9; Indels 4; Gaps 3;

QY 1479 gatgaattgaaatgctgaatgcaacattgtcacaataagcccgacacgtctctgt 1538
DB 1 GATGAAGATTGAAATGTGAGTGCACACTTGTCAACATAAGCCCGACACGCTCTGCT 60
QY 1539 actgaactgtgtgagggcacatttgggcagctgtgcgtcattacggagaccaggtgga 1598
DB 61 ACTGACACTGTGTGAGGGCACGTTTGGGCAGCTGTGCCGTCAATACGGAGACCCAGGTGGA 120
QY 1599 cagggtcctgggacacctggtgtgtgtgtgtccacacctgcacgacatcacacac 1658
DB 121 CAGGTCCTGGCCACCCCTGGCTGCTGTGTGTGTCCACCTGCACGCACATCACCACAC 180
QY 1659 gggcttgcaaatatcttctgtcagagagaacgccttggcattttgggaaagccgt 1718
DB 181 GGGCTTGCAAGATATCTTGTCTCAGAGAGAACGCGCTTGGCATCTTTGGGAAAGCCGCT 240
QY 1719 tcacaccttggctgt 1778
DB 241 TCACCCCTTTGCTGGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT 300
QY 1779 ccagtgcaggagggtcctgcacacatcagatgatctctgcacaaatgccttcagggaag 1838
DB 301 CCAGTGCACAGGAGTCTGTCACACACATCAGTATGATTCCTGCCAATATGCTTCAGGAAG 360
QY 1839 ggcctgagatctcagctcctgcagtggaagattgatcagctgtgtgtgtgtgtgtgtgt 1898
DB 361 GGCCTGAGATCTCCAGTCTGCAGTGGAAAGATTGATCAGTTCGCTGTTCGGAACATGTGA 420
QY 1899 ttgtgaagaatttcagacctgtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1958

Db 421 -TTGGAAGAGTTTCAGACCTGTCTGTGTGGGCGACATGCCATGCGCTGTGCGCT 479
QY 1959 ggtgcacacctgtggtgaaagtgtctattccgggggacaccatccctgcagagctct 2018
Db 480 GGTGCACACCTCTGTGCTGAAAGTGTCTATTCCGGGGACACCATGCCCTGGCAGGCTCT 539
QY 2019 ggtcggatgggaaagatgcacacctctctgatacatgaagccacacctggaagatggtt 2078
Db 540 GGTCCGGATGGGAAAGATGCCACCTCTCTGATACATGAAGCCACCTGGAGATGTTT 599
QY 2079 ggaagagaaagraglggaaaaacacacacagcaacagctcccaagccatcagcgtggggt 2138
Db 600 GGAAGAGGAAGCAGTGGAAAGACACACACAGCAACAGCTCCCAAGCCATCAGCGTGGGAT 659
QY 2139 gcggatgaacgcggagttcattatgctgaacaccttcagccagcgtatgccaaagtccc 2198
Db 660 GCGGATGAACGCGGAGTTCATTATGCTGAACCACTTCAGC--AGGTATGCCAAGT-CC 716
QY 2199 cctcttcagcccaaccttcagcgagaaagtggagttgctt 2240
Db 717 CCTCTTCAGCCCAACTTCAAGCGAGAAAGTGGAGTTGCTTT 758

RESULT 2
BE619259 823 bp mRNA EST 24-AUG-2000
LOCUS 601473130F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 5'
DEFINITION mRNA sequence.
ACCESSION BE619259
VERSION BE619259.1 GI:9890197
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 823)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM623 row: k column: 08
High quality sequence stop: 695.

FEATURES
Source
1..823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876223"
/clone.lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."

BASE COUNT 185 a 219 c 259 g 160 t
ORIGIN

Query Match 27.4%; Score 679; DB 110; Length 823;
Best Local Similarity 92.7%; Pred. No. 4.7e-167;
Matches 757; Conservative 0; Mismatches 55; Indels 5; Gaps 4;

QY 1562 ttggcagctgtgcctcattacggagaccaggtgacaggttcctgggaccctggctg 1621


```

QY 1282 cgtccagagagagtgccagagagatgccattattacttgaatcctgaggaattcata 1341
Db 609 COTCCAGAGGAGTGGCAGAGGATGCCATTATTTACTTGCATCTCCCTGAGGAAATTCATA 668
QY 1342 gttgagggcgtcagcttcccaactccagcagagcgtgacagaggtacagagagagtgccg 1401
Db 669 GTTAGGCGCTG-AGCTTCCCAACTTCCAGCAGAGCGTGCAGAGGTACAGAGGAGTGCG 727
QY 1402 caggacggccccagccccag 1420
Db 728 CAGAGGCCAGCCAGCAG 746

RESULT 4
LOCUS BE260495 676 bp mRNA EST 13-JUL-2000
DEFINITION 601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
mrna sequence.
ACCESSION BE260495
VERSION BE260495.1 GI:9131807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 676)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LUCM176 row: d column: 01
High quality sequence stop: 672.
FEATURES
Location/Qualifiers
1..676
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3503184"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dt priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: This is a NIH_MGC Library."
BASE COUNT 154 a 207 c 176 g 139 t
ORIGIN

```

```

Query Match 27.2% Score 676; DB 106; Length 676;
Best Local Similarity 100.0%; Pred. No. 2.7e-166;
Matches 676; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1093 accagcacttgctgaatgagaaactgctcagttcacaaacttcgagccacaag 1152
Db 1 ACCCAGCACTTGCTGTAATGAGAACTGTGCTTCAACAACTTCGAGCCACAAG 60
QY 1153 attcaaacccagctcaactctaccacccgagacatcttccccctgctcaccagtttccgc 1212
Db 61 ATTCAAAACCCAGCTCAACTCATCCACCCGGACATCTTCCCCCTGCTCACCAGTTCCCG 120
QY 1213 tgtaagaagaggggccccaccctcagtgtagccatggttcagggtgaatgcctctcaag 1272

```

```

Db 121 TGTAAAGAGAGGGCCCCCACCCTCAGTGTGCCCATGGTTCAAGGTGAATGCCCTCCTCAAG 180
QY 1273 taccagctcctccagagggagtggtgcagagaggtgcccattattacttgcaatcctgag 1332
Db 181 TACCAGCTCCGTCCAGAGGAGGTGGCAGAGGATGCCATTATTTACTTGCATCTCTGAG 240
QY 1333 gaattcattagttagggcgtgcagcttcccaacttcacagcagagcgtgcaggaggtacag 1392
Db 241 GAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAGG 300
QY 1393 aggagtgccagagcggccccagccccagcagagaaagaagtcagtcaccagaaaatc 1452
Db 301 AGGAGTGCAGAGCGCGCCAGCCCGCAGCAGAGAAAGAAGTCAGTACCCAGAAATCATC 360
QY 1453 ttcccttggaacagggtcctgccatcccgatgaagattcgaatgtcagtgccacactgtgc 1512
Db 361 TTCTCTTGGAACAGGGTCTGCCATCCCGATGAAGATTTCGAANTGTCACTGCGCACACTTGT 420
QY 1513 aacataagccccgacacacgtctctgtctacttgagctgtggtgaggggcacatttgggcagctg 1572
Db 421 AACATAAGCCCCGACACAGCTCTCTGTCTACTGGACTGTGGTGAGGGCACATTTGGGCAGCTG 480
QY 1573 tgccgtcattacggagacacagtggtgacaggggtcctctgggcccacctggctgtgtgtgtg 1632
Db 481 TGCCCGTCAATTAGCGAGACAGGTGGACAGGGTCTCTGGCACCCCTGGCTGCTGTGTGTGTG 540
QY 1633 tccacctcagcagcagatcacacacagggcttgccagtatcttctgcagagagaacgc 1692
Db 541 TCCACCTCAGCGCATCACCACAGGGCTTGCCCAAGTATCTTGTCTGCAGAGAGAACGC 600
QY 1693 gctctggcatctttgggaaagccgtctcacccttctgctggtggtgtgtgtgtgtgtgtgt 1752
Db 601 GCCTTGGCATCTTTGGGAAAGCGCTTCAACCTTTGCTGTGTGTGTGTGTGTGTGTGTGTGT 660
QY 1753 aaagccttgctcagc 1768
Db 661 AAAGCCTGGCTCCAGC 676

RESULT 5
BE867512 812 bp mRNA EST 27-SEP-2000
LOCUS 601443010F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847226 5',
mrna sequence.
ACCESSION BE867512
VERSION BE867512.1 GI:10316288
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 812)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM548 row: c column: 03
High quality sequence stop: 686.
FEATURES
Location/Qualifiers
1..812
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3847226"

```

/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 200 a 216 c 220 g 176 t
ORIGIN

Query Match 26.7%; Score 662.6; DB 136; Length 812;
Best Local Similarity 97.4%; Pred. No. 9.4e-163;
Matches 716; Conservative 0; Mismatches 14; Indels 5; Gaps 4;

Qy 724 gctttcattctgaagcttcacttaagagagaaactcttgggtcctcaaaagcaagag 783
Db 1 GCTTTCATCTGAAGCTTCACCTTAAGAGAGAAACTTCTTGGTGCTCAAGCAAGGAG 60
Qy 784 atg99cctccagttg99acagctgcctcctccatctatgctgtgtcaagagacggg 843
Db 61 ATGGGCTCCCAAGCTTGGGACAGCTGCCATCGCTCCCATCATTTGCTGTCAAGGACGG 120
Qy 844 aaagcatcactcatgaaggaagagatttggctgaagagctgtgtactcctccagat 903
Db 121 AAAGCATCATCTATGAAGGAGAGAGATTGGCTGAAGAGCTGTGTACTCTCCAGAT 180
Qy 904 cctggtgctgttttgggtgtagaattccagatgaagcttcattcaaccatctgt 963
Db 181 CCGTGTCTGC-TTTGTGGTGAAGATGCCAGATGAAGCTTCATTCACCCATCTGT 239
Qy 964 gagaatgccactttcaagagtagacaaggaagcagatgccccgtggtcctgtgtgtt -1023
Db 240 GAGAATGCCACTTTTCAGAGGTACCAAGAAAGGAGAGTGCCTGCTGTGGTGT 299
Qy 1024 cacatggccagcatctgtctgttgacagcaggtaccagagtgatgagagagttt 1083
Db 300 CACATGGCCCCAGCATCTGTCTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTT 359
Qy 1084 gggcctgacacccagcacttggctgtaagtgagaaactgtgctcagttcacaaccttcgc 1143
Db 360 GGGCTGACACCCAGCATTGGTCTGAATGAGAACTGTGCTCAGTTCACACCTTCGC 419
Qy 1144 agccacaagattcaaacccagctcaacctcaaccctccagcatcttccccctgtcacc 1203
Db 420 AGCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCTGTCTACC 479
Qy 1204 agtttcgctataaagagagagcccccactcaagtgtgcccattggttcagggtgaatgc 1263
Db 480 AGTTTCGCTGTAAAGAGAGGGCCCCACCTCTAGTGTGCCCATGGTTTCAGGGTGAATGC 539
Qy 1264 ctctcaagtaccagctccgtccagagagg--agtggcagaggatgccattattact- 1320
Db 540 CTCCTCAAGTACCAGCTCCGTCACAGGAGGCGAGTGGGACAGGGATGCCATTATTACTC 599
Qy 1321 tgcatactcaggaattcaatagtgagcgtcagcttcccaactccagcagagagtg 1380
Db 600 TGCAATCTGAGGAATTCATATTGAGCGCTGCAGCTTCCCAACTT-CAGCAGAGCGTG 658
Qy 1381 cagagtagcagagagtagtcagagacgcccagcccccagcagagaaagagcagtagtac 1440
Db 659 CAGGAGTACAGGAGAGTGGCCAGGAGCGGCGCCAGCCAGCAGAGAAAGAAAGTCAGTAC 718
Qy 1441 ccagaaatcatcttc 1455
Db 719 CCAGAAATCATCTCC 733

RESULT 6
BE902696 735 bp mRNA EST 29-SEP-2000
LOCUS 601677393f1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926.5',
DEFINITION mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE902696
BE902696.1 GI:10393148
EST.
human.
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LICM841 row: j column: 23
High quality sequence stop: 732.
Location/Qualifiers
1..735

FEATURES
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3959926"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGCAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 168 a 206 c 203 g 157 t 1 others
ORIGIN

Query Match 26.4%; Score 655.6; DB 137; Length 735;
Best Local Similarity 97.4%; Pred. No. 6.2e-161;
Matches 719; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

Qy 1190 tccccctgtccacagtttccgtgtaagaaggaggggccccaccctcagtgcccatgg 1249
Db 1 TCCCCCTGCTCACCAGTTTCCCGCTGTAAAGAGAGGGGCCACCTCAGTGTGCCATGG 60
Qy 1250 ttcagggtgaatgcctcctcaagtaccagctccgtccccagaggagtgaggagggatg 1309
Db 61 TTCAGGATGAATGCTCTCTCAAGTACCAGCTCCGTCCTCCAGGAGGAGTGGCAGAGGATG 120
Qy 1310 ccattattactt-gcaatcctcaggaaattcatagttgagcgctgcagcttcccaacttc 1368
Db 121 CCATTATTACTTNGCAATCCCTCAGGAATTCATAGTTGAGCGCTGCAGCTTCCCAACTTC 180
Qy 1369 cagcagagcgtgcagaggtacagagagagtgcgagagagcggccccagccagcagagagaaa 1428
Db 181 CAGCAGAGCGTGCAGGAGTACAGGAGGAGTGCAGGAGGCGGCCAGCCCGCCAGCAGAGAAA 240
Qy 1429 aagaagtcagtcaccagaaatcatcttcttggaacaggtctgcctccatccgatgaagatt 1488
Db 241 AGAAGTCAGTACCAGAAATCATCTTCTTGGACAGGGTCTGTCATCCCATGAAGATT 300
Qy 1489 cgaaatgicagtgccacacttgtcaacataagccccacacgtctctctgtactgtgactgt 1548
Db 301 CGAAATGTCAAGTGCACACTTGTCAACATAAGCCCGACACAGTCTCTCTACTTGGACTGT 360
Qy 1549 ggtgagggcacatttggcagctgtgccgttcattacagagaccaggtggacagaggtcctg 1608
Db 361 GGTGAGGCGACATTTGGCGAGCTGTGGCGTCATTACCGAGACACAGGTGGAGGAGGTCCTG 420

Tissue Procurement: DCTD/DTP
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 plate: LLC524 row: 9 column: 18
 High quality sequence stop: 668.

FEATURES

Location/Qualifiers
 1. 949
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3838121"
 /clone_lib="NIH_MGC_9"
 /tissue_type="adenocarcinoma cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 225 a 263 c 281 g 179 t 1 others
 ORIGIN

Query Match 26.2%; Score 649; DB 135; Length 949;
 Best Local Similarity 92.6%; Pred. No. 3.6e-159;
 Matches 773; Conservative 0; Mismatches 45; Indels 17; Gaps 8;

Qy 672 tccacatggtgttagccagagaagagggtcagggaacttccctggtcgtagcttcat 731
 Db 1 TCCACATGGTGTAGCCAGAGAGAGGGGTcAGGGACTTCCCTGCTGTAGCTTCAT 60

Qy 732 ctgtaagttcacttaagaagagaactctgtgtctcaagaagaagagatggcct 791
 Db 61 CTGTAAGCTTCACTTAAGAGAGGAACCTCTTGGTGCTCAAAAGCAAGAGATGGCCT 120

Qy 792 cccagttggagacagctgccatcgctccatcattgctgctcaagacgggaaagcat 851
 Db 121 CCCAGTTGGACAGCTGCCATCGCTCCCATCATTTGCTGTCAGGACGGGAACCAT 180

Qy 852 cactcatgaaggagagagattttggctgaagagctgttactctccagatccctgtgc 911
 Db 181 CACTCATGAAGGAAGAGAGA-TTTGGCTGAAGAGCTGTACTCTCCAGATCCTGGTGC 239

Qy 912 tgcctttgtgtgtagaattccagatgaagcttcaatcaaccatctgtgagaatgc 971
 Db 240 TGC-TTTGTGGTGTAGAAATGTCAGATGAAGCTTCATTCAACCCCATCTGTGAGATGC 298

Qy 972 cacccttcagaggtaccagaagaggcagatgcccccgctggccttgggtcacatggc 1031
 Db 299 CACCTTTCAGAGGTACCAAGGAAGGCAGATGCCCCGCTGGCTTGGTGTTCATATGC 358

Qy 1032 cccagcatctgtctgttgacagcaggtaccagcagctgagatggagaggtttgggcctga 1091
 Db 359 CCCAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGGTTTGGGCCCTGA 418

Qy 1092 caccagcaacttggctcctgaatgagaactgtgctcagttcaacaaccttcgcagccaca 1151
 Db 419 CACCCAGCACTTGGTCTGTAATGAGAACTGTGCTCAGTTCAACAACCTTCGCAGCCACA 478

Qy 1152 gattcaaacccagctcaactcatccaccggagcatcttccctctgtcaccagtttccg 1211
 Db 479 GATTCAAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCTCTGCTCACCAGTTTCCG 538

Qy 1212 ctgttaagaagaggggccccaccctcagtgccccatggtttcagaggtgaatgcctctcaa 1271
 Db 539 CTGTGAAGAAGAGGGGGCCCCACCCTCAGTGTGCCATGGTTCAGGGTGAATGCCCTCTCAA 598

Qy 1272 gtaccagctccgtcccccaggaggaggtggcagagggatgccattattac-ttgcaactctg 1330

Db 599 GTACACGCTCGT-CCAGGAGGGAGTGGCAGAGGATGCCATTATTCTTTGCAATCCTG 657
 Qy 1331 aggaattcatagttgagcgtgcagc-----ttcccaactccagca-gagcgtgcag 1383
 Db 658 AGGAATTCAATAGCTTGAGGGCGGTGCAAGGCTTCCCAAACTCCAGCAGAGGCTGCAG 717
 Qy 1384 ---gagtcacagggagtgatgcgaggacgcccagcccccag---cagagaaagaagtcag 1437
 Db 718 GGAGTACCCGAGGAGTGGCCGAGCAGCGCCAGCCAGCCAGCAGAGAAAGTCAAGTCAG 777
 Qy 1438 taccagaaatcatcttcccttggaacaggggtctgcatcccgatgaagattcgaa 1492
 Db 778 GCACCAGAAATCACTTCTCTGGAACAGGCTTGGCACCCTCCGATGAAGATCGAAA 832

RESULT 9

BE747163 992 bp mRNA EST 15-SEP-2000
 LOCUS 601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
 DEFINITION mRNA sequence.

ACCESSION BE747163
 VERSION BE747163.1 GI:10161155

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Tel: (301) 496-1550

Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

plate: LLC524 row: 1 column: 14

High quality sequence stop: 781.

FEATURES

source

Location/Qualifiers
 1. 992

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3838237"

/clone_lib="NIH_MGC_9"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Size-selected >500bp for average
 insert size 1.8kb. Library constructed by Ling Hong in
 the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 225 a 287 c 313 g 167 t

ORIGIN

Query Match 25.9%; Score 641.6; DB 135; Length 992;
 Best Local Similarity 97.9%; Pred. No. 3.2e-157;
 Matches 693; Conservative 0; Mismatches 9; Indels 6; Gaps 4;

Qy 991 ggaagcagatgcccccgtggccttggtgttcaatgcccccagcatctgtgtctgtg 1050
 Db 1 GGAAGCAGATGCCCCGCTGGCC-TGGTGGTTTCACATGGCCCGCAGCATCTGTCTGTG 59

Qy 1051 gacagcaggtaccagcagtgatggagaggtttggcctgacacccagcattgtctgt 1110

Matches	637;	Conservative	0;	Mismatches	7;	Indels	2;	Gaps	2
---------	------	--------------	----	------------	----	--------	----	------	---

QY	1475	tccgatgaagatcgaatactgagtgcacacttgtcaacataagaagccccacagctctc	1534
Db	1	TCCCGATGAAGATTGCAAAATGTCAAGTGCACACTTGTCACATTAAGCCCCGACAGCTCTC	60
QY	1535	tgctactggacctgtggttgaggaccattttggcgagctgtgcgcgttaattacagagaccagg	1594
Db	61	TGCTACTGGACTGTGGTAGGGCACATTTGGGACAGCTGCCGTCTATTACGAGACCAGG	120
QY	1595	tggacaaggctctgggcaacctggctgctgtgttgtgttcaccactgcacagataacc	1654
Db	121	TGGACAGGCTCTGGGCACCTGGCTGTGTGTGTGTGTCACCTGCACGCAGATACCC	180
QY	1655	acacgggcttgccaagtatcttctgcagagagaacgccttggcatcttttgggaaagc	1714
Db	181	ACACGGGCTTGCCAAGTATCTTGTGCAGAGAGAAGCGCCCTTGSCATCTTTGGGAAGC	240
QY	1715	cgttcaacctttgctggtggtgttgcgcccaaccagctcaaacgctggtcccagcagtagcc	1774
Db	241	CGCTTTCACCCCTTGTGTGTGTGTGTCGCCCAACACAGCTCAAAGCCTGGCTCCAGCAGTACC	300
QY	1775	aaaaccagtcgacagggtctctgcacacatcatgatatgtattccgtgccaatgccttcagg	1834
Db	301	ACAACACAGTCGCAGGAGGTCTCTGCACCAACATCAGTATGATTTCTTGCCAAAATGCCCTTCAGG	360
QY	1835	aagggctggaatctccagctcctcagtggaagatgtagatcagttcgctgttgcaaat	1894
Db	361	AAGGGCTGAGATCTCCAGTCTCGACTGGGAAGATTGATCAGTTCGCTGTTGGAACAT	420
QY	1895	gtgatttggaaagatttcagacctgtgtgtgcgcgactgcaagcatcgcttttggctgtg	1954
Db	421	GTGATTTGGAAGAGTTTTAGAACCTGTCTGTGTGCGGACTGCAAGCATGCGTTTGGCTGTG	480
QY	1955	cgctgggtcacacctgtgctggaaagtgtgttatctccggggagacacatgccttcgcgagg	2014
Db	481	CGCTGGTGCACACTCTGGCTGGGAAGTGTGCTATTTCGGGGACACCAATGCCCTGCGAGG	540
QY	2015	ctctggctcgagatggggaagatgccacctctctgatatcatgaagccaccctggaagatg	2074
Db	541	CTCTGGTCCGATGGGGNMAAGATGCCACCTCTCTGTATACATGAAGCCA-CTTGGGAAGATG	599
QY	2075	gtttggaagagaagcagtggtgaaaaagacacacagacacacagctcccaa	2120
Db	600	GTTTGGCAGAGGAAGCCGTGG-AAAGACACACAGCACACGTCCTCCCAA	644

RESULT	13
BE892893	
LOCUS	BE892893 790 bp mRNA EST 29-SEP-2000
DEFINITION	60143573BF1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3920792 5', mRNA sequence.
ACCESSION	BE892893
VERSION	BE892893.1 GI:10353525
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	NH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE	1 (bases 1 to 790)
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert_Strausberg@nih.gov Tissue Procurement: ATCC/DCTD/DTF CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCW739	row: 1	column: 09
High quality sequence stop: 662.		
FEATURES	Location/Qualifiers	
source	1..790	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:3920792"	
	/clone_lib="NTH_MGC_72"	
	/tissue_type="melanotic mel"	
	/lab_host="DH10B (phage-res"	
	/note="Organ: skin; Vector:	
	Size2: SalI; Cloned unidir	
	Average insert size 2 kb.	
	Technologies."	
BASE COUNT	185 a 221 c 220 g 164 t	
ORIGIN		
Query Match	24.6%;	Score 610.4;
Best Local Similarity	93.6%;	Pred. No. 4.5e
Matches 726;	Conservative 0;	Mismatches
QY 903	tctcgtgtcgtcttttgggtggtgagaatgtctccagatg	
Db 1	TCTCGTGTCTGCTTTTGTGGTGTGAATGTCCTCAGATG	
QY 963	taagaatgccacctttcagaggtaccagaaagagcag	
Db 61	TGAGATGCCACCTTTCAGAGGTACCAAGAAAGGCAG	
QY 1023	tcacatggccccagcatctgtctgttggtggacagcaggt	
Db 121	TCACATGGCCCCAGCATCTGTGCTTGTGGACAGCAGGT	
QY 1083	tgggcctgacacccagcagcacttggctcgtgaatgagaact	
Db 181	TGGGCGCTGACACCCAGCAGCTTGGTCTCTGAATGAGAACT	
QY 1143	cagccacaagatccaacccagctcaacactcatccacc	
Db 241	CAGCCACAAGATTCAACCCAGCTCAACCTCATPCCACCC	
QY 1203	cagtttcgcgtgtaagaaggagggtcccccaccctcagtg	
Db 301	CAGTTTCCGCTGTAAGAGAGGGGCCCCACCCTCAGTG	
QY 1262	gcctctcaagtaccagctccgtccccagagagggagtgg	
Db 361	GCCTCTCAAGTACCAGCTCCGTCCAGGAGGGAGTGGG	
QY 1322	gcaatcctgaggaattcatagttgaggcgctgcagctttt	
Db 421	GCAATCTGAGGAATTATAGTTTGAGGGCGCTGCAGCTTT	
QY 1381	caggaggtacaggagaggtggtcaggagcaggtccagcccc	
Db 481	CAGGAGTACAGGAGGAGTGCCGAGACGCGCCACGCCCA	
QY 1440	cccagaaatcatctctcttggaaacagggtctgcacatccc	
Db 541	CCCAGAAATCATCTTCCTTGGAAACAGGGTCTGCCATCCC	
QY 1498	agtgcacacattg---tcaacataaagccccgcagacagtc	
Db 601	AGTGCCACACTTGGTCTCATATAAAGCCCCCGAACACGTC	
QY 1555	ggcacattt-gggcagctgtgcctgcattaccgg-aagc	
Db 661	GGCACATTTGGGGCAAGTGTGGCGGCATTTACGGGAAGAC	
QY 1613	ccctggctgct---gtgtttgtgttcccacactgcagcgc	
Db 721	CACCTGGGCTTGGTGGGTTTGGGTTTCCACCGCGCAGC	

ORIGIN

Query Match 23.3%; Score 578; DB 135; Length 940;
Best Local Similarity 89.5%; Pred. No. 1.5e-140;
Matches 725; Conservative 0; Mismatches 70; Indels 15; Gaps 9;

QY 672 tccacatggtgttagccagagaaaggggtcagggaactctccctgggtgtagcttcat 731
DB 1 TCACATGGTGTGTAGCCAGAGAGAGGGGTCA-GGACTCTTCCCTGGTGTAGCTTTTCAT 59

QY 732 ctgtaagcttcaactaaagagagaaactcttggctcacaagcaaaagagatggcct 791
DB 60 CTGTAAGCTTCACTTAAAGAGAGGAAACTTCTTGGTGTCAAAAGCAAGAGATGGCCCT 119

QY 792 cccagttgggacagctgccatcgctcccaatcattgctgctcaaggacgggaaaaagcat 851
DB 120 CCCAGTTGGGACAGCTGCCATCGCTCCCATCATTTGCTGTGTCAGGACGGGAAAAGCAT 179

QY 852 caactcatgaaggaagagagattttggctgaagagctgtgtactcctccagatccctgggtgc 911
DB 180 CACTCATGAAGGAAGAGAGATTTTGGCTGAAGAGCTGTGTACTCTCCAGATCCTGGTGC 239

QY 912 tgcttttggtggttagaatgccagatgaagcttcatcaccatctgtgaaatgc 971
DB 240 TGCTTTTGTGGTGTAGAAATGTCCAGATGAAGCTTCAACCCATCTGTGAGAAATGC 299

QY 972 caccittcagaggtaccaggaaggaagcagatgcccccggtggccttgggttcacatgac 1031
DB 300 CACCTTTTCAGAGGTACCAAGGAAGGACAGATCCCCCGTGGCTTGGTGTTCACATGG- 358

QY 1032 cccagcatctgtctgttgagcagcaggtaccagcagtgatggagaggtttgggcctga 1091
DB 359 CCCAGCATCTGTGCTTGTGGAGCAGCAGGTACCAGCAGTGGATGGAGAGG-TTGGGGCTGA 417

QY 1092 caccagcactggtcctgaatgaatgaaactgtcctcagttcacacacttcgcagccacaa 1151
DB 418 CACCAGCAGCACTTGGTCTGAATGAGAACTGTGCTCAGTTACAA-CTTCGCAGCCACAA 476

QY 1152 gattcaaacccagctcaactcatccacccggacatcttccccctgctcaccagtttccg 1211
DB 477 GATTCAAAACCCAGCTCAACCTCATCCA-CCGGACATCTTCCCTGTGCTCACCAGTTTCGG 535

QY 1212 ctgtaagaagggggcccccacactcagtggtcccatggttcaggggtgaatgctcctcaa 1271
DB 536 CTGTAAGAAGAGAGGG-CCCAACCTCAGTGTGCCATGTTTCAAGGTGAATGCTCCTCAA 594

QY 1272 gtaccagctccgtcccagagagagtgccagaggaatgcc-----attattactgca 1324
DB 595 GTACCAGTCCGTCCCGAGAGGGAGTGGCAGAGGGATGCCCATTTATATTGACAATC 654

QY 1325 atccctgaggaattcatagttgaggcgctgcagcttcccacacttcca-gcagagcgtgcag 1383
DB 655 CCTGCACGGAATCCATAGATAGGCGCGAGAGATACCAAGATACGAGCGAGGCGGCGAG 714

QY 1384 gagtacagagagagtcgcagcagcagggccagccccagcagagaaagagagtcagttacca 1443
DB 715 GAGTCCAGAGGAAGTGCACGAGGGGGCCAG-CCCAGGAGCGGAAAGAGCGTGAAGACCCA 773

QY 1444 gaaatcatcttcttgggaacaggggtctgcc 1473
DB 774 GAATACAGCTCAGGAAAGGGGCGCAACC 803

Search completed: February 18, 2001, 07:27:13
Job time: 22508 sec

OM of: US-09-434-382-2 to: GenEmbl.* out_format : pfs
Date: Feb 18, 2001 7:51 PM
About: Results were produced by the GenCore software, version 4.5.
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL-frame+ p2n.model -DEV-xlp
-Q=/cgnl_1/USPTO_spool/US09434382/runat_16022001_105748_29682/app_query.fasta_1.895
-DB=GenEmbl -QPM=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX-blosom62 -TRANS=human4.0.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=10 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM-ext MINLEN=0
-MAXLEN=200000000 -USER=US09434382.cgnl_1.7821 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search Information block:

Query: US-09-434-382-2
Query length: 826
Database: GenEmbl.*
Database sequences: 1118133
Database length: -1736092196
Search time (sec): 2992.230000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
gb_pr4:AK001392	+ 4266.50	4573.21	1.4e-246	2976	AK001392 Homo sapiens cDNA FLJ
gb_in2:AF215894	+ 1144.00	1220.95	7.5e-60	2690	AF215894 Drosophila melanogast
gb_in1:AC007352	+ 1136.00	1174.61	2.9e-57	190574	AC007352 Drosophila melanoga
gb_in1:AC003830	+ 1136.00	1171.79	4.1e-57	261846	AE003830 Drosophila melanoga
gb_in6:AC017383	+ 1132.50	1175.26	2.6e-57	155873	AC017383 Drosophila melanoga
gb_in9:AC007417	+ 1128.00	1169.72	3.3e-57	125448	AC007417 Drosophila melanoga
gb_in1:AC005277	+ 1012.00	1045.64	4.4e-50	118788	AC005277 Homo sapiens chromo
gb_in1:SPAC1D4	+ 785.50	814.25	3.4e-37	31201	269239 S.pombe chromosome I c
gb_in1:AC022354	+ 677.50	694.77	1.5e-30	46335	AC022354 Arabidopsis thaliana
gb_in1:AR023046	+ 647.50	658.25	1.6e-28	75289	AR023046 Arabidopsis thaliana
gb_in1:SCYK8079C	+ 613.00	647.94	6.2e-28	3685	228304 S.cerevisiae chromosome
gb_in1:AF188714	+ 610.00	630.11	6.1e-27	19158	AF188714 Emeritella nidulans
gb_in2:CELE04A4	+ 591.50	608.68	9.5e-26	22846	AF038611 Caenorhabditis elega
gb_in1:AC006719	+ 591.50	589.05	1.2e-24	209365	AC006719 Caenorhabditis eleg
gb_in1:AP001550	+ 487.50	480.74	1.3e-18	143209	AP001550 Oryza sativa genom
gb_in1:SPBC3D6	+ 430.50	432.72	6.0e-16	33229	295620 S.pombe chromosome II
gb_in1:AE000936	+ 300.00	301.48	1.2e-08	1836	AE000936 Methanobacterium the
gb_in1:AE000316	+ 275.00	275.77	3.3e-07	10409	AE000316 Escherichia coli K-1
gb_in2:D90858	+ 275.00	273.44	4.5e-07	13548	D90858 E.coli genomic DNA, KC
gb_in2:D90857	+ 275.00	272.14	5.3e-07	15676	D90857 E.coli genomic DNA, KC
gb_in2:ECU58768	+ 273.00	281.59	1.6e-07	4235	U58768 Escherichia coli ela
gb_in2:AP001512	+ 262.50	232.55	8.5e-05	300550	AP001512 Bacillus halodurans
gb_in2:AP000005	+ 254.00	228.37	0.0001	172000	AP000005 Pyrococcus horikosh
gb_in2:U67591	+ 246.50	245.67	1.6e-05	9829	U67591 Methanococcus jannasch
gb_in1:AF209151	+ 245.50	257.99	3.3e-06	2169	AB029151 Homo sapiens D29 mRN
gb_in2:SYCCPCNC	+ 243.50	221.40	0.0004	105795	D64001 Synchocystis sp. PCC
gb_in1:AF270015	+ 241.50	247.97	9.3e-06	3376	AF270015 Staphylococcus epide
gb_in1:AE005109	+ 241.00	237.98	4.2e-05	12026	AE005109 Halobacterium sp. NR
gb_in1:AK024822	+ 237.50	253.37	5.9e-06	1386	AK024822 Homo sapiens cDNA: FI
gb_in1:AE0002234	+ 233.00	232.71	8.3e-05	8268	AE0002234 Chlamydia pneumonia
gb_in1:AE001588	+ 233.00	229.46	0.0001	11933	AE001588 Chlamydia pneumonia
em_bal:AP002545	+ 233.00	200.90	0.0049	298650	AP002545 Chlamydia muridarum
gb_in2:CNBPAX04	+ 231.00	199.09	0.0062	288440	AJ248286 Pyrococcus abyssi c
gb_in2:BSUB0013	+ 228.50	198.86	0.0064	218470	Z99116 Bacillus subtilis com
gb_in2:BAJUH642	+ 228.50	196.58	0.0086	282700	D84432 Bacillus subtilis DNA
gb_in2:AE001039	+ 226.50	216.46	0.0007	23533	AE001039 Archaeoglobus fulgid
gb_in1:AE002331	+ 225.00	221.53	0.0003	11765	AE002331 Chlamydia muridarum
gb_in2:PGFAGAGEN	+ 225.00	227.49	0.0002	5653	X95938 P.gingivalis rnhB & pga
gb_in1:AE001308	+ 224.50	220.56	0.0004	11627	AE001308 Chlamydia trachomat
gb_in2:TPU93844	+ 221.00	228.46	0.0000	3120	U93844 Treponema pallidum lipd
gb_in1:AE001252	+ 221.00	216.84	0.0006	11583	AE001252 Treponema pallidum s

gb_ba3:SCC77 - 216.00 201.98 0.0043 33801 ! AL136503 Streptomyces coel
gb_ba2:AP000062 + 213.00 180.96 0.0635 251700 ! AP000062 Aeropyrum pernix
gb_htg20:AL158037 + 204.00 172.02 0.1998 231916 ! AL158037 Homo sapiens chr
gb_htg14:AC049145 - 194.00 208.05 0.0020 1184 ! AC049145 Giardia intestinal
seq_name: gb_pr4:AK001392

seq_documentation_block:
LOCUS AK001392 2976 bp mRNA 22-FEB-2000
DEFINITION Homo sapiens cDNA FLJ10530 fis, clone NT2RP2000985.
ACCESSION AK001392
VERSION AK001392.1 GI:7022621
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens teratocarcinoma cell_line:NT2 cDNA to mRNA,
clone_lib:NT2RP2 clone:NT2RP2000985.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (sites)
AUTHORS Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,
Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,
Wagatsuma,M., Hosoiri,T., Kaku,Y., Kodaira,H., Kondo,H.,
Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,
Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,
Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,
Nakamura,Y., Nagahari,K., Masuho,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
Unpublished (2000)

2 (bases 1 to 2976)
Isogai,T. and Otsuki,T.
Direct Submission

TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

Submitted (16-FEB-2000) to the DDBJ/EMBL/Genbank databases. Takao
Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,
Tel.81-438-52-3951, Fax:81-438-52-3952)

COMMENT
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing and clone selection:
Helix Research Institute (supported by Japan Key Technology Center
etc.) and Department of Virology, Institute of Medical Science,
University of Tokyo.

FEATURES
source

1..2976
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone="NT2RP2000985"
/note="cloning vector: pME18SFL3; mRNA from NT2 neuronal
precursor cells after 2-weeks retinoic acid (RA)
induction."
75..2639

CDS

/codon="unnamed protein product"
/note="start=1"
/protein_id="BAA91666.1"
/db_xref="GI:7022622"
/translation="MVALCSLRAGRTMSQRTISQAPREERKDPRLHLRTRE
KRGSCGCGPNTVYLVQVVAAGSRDSCAALYVSEFNRYLFCNGEVRQLRQDKLKY
ALNDLFLTRMWSNVGLTGLKPCVLGSPQLEKYLEIAIKIFSGPLKG
ENEPHLPGVSQRRGVRDSTVYQIPHSQRGKHQWQSPRLSPRLSPERSDSES
SLGLKPLPHVQNLKAWLOQYVQCCVLEHLSIMPAKLOEAGETSSPAVERLI
AVKDGSIPTHEGRIELBELCTPPDGAAPVVCPEDEFIOPICNATFORVQGNAD
AVYLVHMAPASVLDSYVQOMMEREGPTQHLVNLNENCAVSHNRSHKIQTLNLI
HDPIDPLLSIFRCKEPTLSVPMVQGECLTKQLRPREWRDRAITCPEEPIVAL
QUPNFQSQVQYRRAQDGPAPAKRSQYPEIFLGTGSAIPKIRNVATLVNISPD
TSLLDGCGTGCQCRHYGQDQVRVLGTAAVFSHLHADHHTGTSVLLQALRALA
SLGLKPLPHVQNLKAWLOQYVQCCVLEHLSIMPAKLOEAGETSSPAVERLI
SLLRPTCDLEEFQTCVLRCHKAFCGALVHTGKVVYSGDTMCALVRMGKDTLL
IHEATLDGLEAEVKEKTHISTSOALISVGRNNAEIMLNFHSQYAKVPLSPNSE
KVGAVFDHMKVCLGDPPTPKLIPPLKALFAGDIEEMERREKREKQVRAALLSHEL
AGLEDGSLRSRSGTQRSRRRSEPSDGLDPELRRLVCVCTHPAPVSALLAGRS"

37 euArgHisLeuArgThrArgGluLysArgGlyProSerGly..... 50

37 euArgHisLeuArgThrArgGluLysArgGlyProSerGly..... 50


```
614 AlaGluIleSerSerPro.....AlaValGluArgLeuIleSe 626
      : : : : :
1834 GAACCTCTTGGCTAGTCCCTGAGCGGTGAACAAGTTTCAACGCTCTA..... 1878
      : : : : :
626 rSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuValA 643
      : : : : :
1879 .....GGATTACGTCCCATATCCACCTGCTAGTTA 1909
      : : : : :
643 rHisCysLysHisAlaPheGlyCysAlaLeuVal..... 654
      : : : : :
1910 GGCACGTGCCCAACTCTTTTCGGAATAAGCCTAACTCTGGCGGCAAGCAC 1959
      : : : : :
655 HisThrSerGlyTrpLysValValTyrSerGlyAspThrMetProCysG 671
      : : : : :
1960 AATAGCGAACCCTGCTCAAGATCTACTTACGCGCGACACTATGCCATGTCA 2009
      : : : : :
671 uAlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAla 688
      : : : : :
2010 GGATCTGATCGATCTGGCGCGGACTCCACCGTTCTTATTCAGAGCGGA 2059
      : : : : :
688 hrLeuGluAspGlyLeuGluGluAlaValGluLysThrHisSerThr 704
      : : : : :
2060 CAATGAGGAGCACTTGGAGGAGGCGCGGTTAAAGAGCGCACACGACC 2109
      : : : : :
705 ThrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMe 721
      : : : : :
2110 GTGTCGACGGCTATTTCAGCAGGGCGGAATATGACGCCGCCATACGAT 2159
      : : : : :
721 tIleuAsnHisPheSerGlnArgTyrAlaLysValPro.....LeuPheSerP 737
      : : : : :
2160 TCTACCCCACTTTTCGACGGTTACGCTAAGTCCCGCGGATTAACCGAGTG 2209
      : : : : :
737 roAsnPheSerGluLysValGlyValAlaPheAspHisMetLysValCys 753
      : : : : :
2210 ATGAGGATATGCAACGAGTGTGAATAGCCTTTGATATATGGAAGTGACC 2259
      : : : : :
754 PheGlyAspPheProThrMetProLysLeuIleProProLeuLysAlaLe 770
      : : : : :
2260 ATTGAGGATCTGCAGCACTACCAACAAGCTCTACCCGCCCTTTTCGCGAT 2309
      : : : : :
770 uPheAlaGlyAspIleGluGluMetGluGluArgArgGluLysArgGluL 787
      : : : : :
2310 GTACGCGGAGTACACGGAAGACTGAGGAGCGGGCTGTCAACGGGGAGC 2359
      : : : : :
787 euArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAla 800
      : : : : :
2360 TGAACACAGGCGCA.....AAGCGAAACTGGCA 2388
      : : : : :
seq_name: gb_htgl.AC007352
```

```
seq_documentation_block:
LOCUS AC007352 190574 bp DNA HTG 24-FEB-2000
DEFINITION Drosophila melanogaster chromosome 2 clone BACR19J16 (D626) RPCI-98
19-J-16 map 47A-47B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 36 unordered pieces.
ACCESSION AC007352
VERSION AC007352.8 GI:7025687
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS Celiniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenihoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K.,
Kearney,L., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D.,
Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
```

Sequencing of Drosophila melanogaster

Unpublished

2 (bases 1 to 190574)

Celiniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenihoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.

Direct Submission

Submitted (22-APR-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Feb 24, 2000 this sequence version replaced gi:7018750.

For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently consists of 36 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 678: contig of 678 bp in length
* 679 758: gap of unknown length
* 759 1358: contig of 600 bp in length
* 1359 1438: gap of unknown length
* 1439 2373: contig of 935 bp in length
* 2374 2453: gap of unknown length
* 2454 2962: contig of 509 bp in length
* 2963 3042: gap of unknown length
* 3043 3889: contig of 847 bp in length
* 3890 3970: contig of 1468 bp in length
* 3971 5517: gap of unknown length
* 5518 6572: contig of 1055 bp in length
* 6573 8652: gap of unknown length
* 8653 8387: contig of 1735 bp in length
* 8388 8467: gap of unknown length
* 8468 9060: contig of 593 bp in length
* 9061 9140: gap of unknown length
* 9141 11167: contig of 2027 bp in length
* 11168 11247: gap of unknown length
* 11248 14518: contig of 3271 bp in length
* 14519 14598: gap of unknown length
* 14599 18937: contig of 4239 bp in length
* 18938 18917: gap of unknown length
* 18918 24324: contig of 5407 bp in length
* 24325 24405: gap of unknown length
* 24406 29984: contig of 5580 bp in length
* 29985 30064: gap of unknown length
* 30065 35794: contig of 5730 bp in length
* 35795 35874: gap of unknown length
* 35875 44808: contig of 8934 bp in length
* 44809 44888: gap of unknown length
* 44889 58222: contig of 13334 bp in length
* 58223 58302: gap of unknown length
* 58303 78322: contig of 20020 bp in length
* 78323 78402: gap of unknown length
* 78403 122449: contig of 43847 bp in length
* 122450 122329: gap of unknown length
* 122330 179102: contig of 56773 bp in length
* 179103 179182: gap of unknown length
* 179183 179407: contig of 225 bp in length
* 179408 179487: gap of unknown length
* 179488 180592: contig of 1105 bp in length
* 180593 180672: gap of unknown length

Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Ibegwam, C., Jatali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, A.T., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarri, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Swirskas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wassarman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 261846)
Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
On Oct 9, 2000 this sequence version replaced gi:7303755.
Location/Qualifiers
1. 261846
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2R"
join(960..1069,1135..1237,1290..1357,1416..1731,2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8315)
/gene="Pfk"
/product="Ct11302"
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
<960..>8315
/gene="Pfk"
/note="CG4001"
/map="46E3-46E6"
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
join(1026..1069,1135..1237,1290..1357,1416..1731,2658..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8059)
/gene="Pfk"
/note="Pfk gene product [alt 1]"
/codon_start=1
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
/protein_id="AAF58840.1"
/db_xref="GI:7303792"
/translation="MHSIKFVPTKLPFLFLEINGRIPICRHPGPTTFRLEISNKTTPPIRQKTFNIGTCVTRSHLCCPRDISGNLTLLSVKNCRRHCICKLRSDGQKNDSPGEKIQDKSAQPCGKPINLNGFLNAYSEKNVKKKSPKPKCKSKSVDELKRCLETRQDVDFHPVFKKGLKAVFTSGGDSQGMNAVRACVMAIYLGCXVYFIREGLYQMDGGCQIOEAWSSVSIHRGTTIGSARQDFRQCRLKAANLTQRTITLVVIGDGLNLFQEWSSLLDELVKNTITTEQEKFNVLHVLGVLSGDSNDPFGDTMTGTDALHRIEIDAISSTAYSHORTFIMEVAGRHCGYLAFIPEMPKVDWDRCSOLAORSAGORLNIVIVAEAGMDRGHPTAEDVKVIDERLKHDAITVLGHVORGGNPSAFDRILACRMGAETLALMEATKDSVRPLMECVRTQAVAKMAEKRWADAVKLRGSRFERNLETYKRLTKLPKPKENADGKLEGRVLAVMHIGAPACMNAAVSRVNAIYRGDVVYIGKLVYIHRGTYREGYQGVGGGAYLGTKRTLPKGKFEIARLKEFKIQGLITIGFESYHAAGQIADQRNYPQFCIPVIVIPSTISNNVPGTFEFLGCDTGLNEITEICDRIRQAQGTKRRVVFETWGGYCYLATLAGAGDAAYIYEKFSIKLOQDVYHMAKMAEVSRLILRKEKASENSTDFLYRLYSEEGGLTCRNNILGHMQQGSPTPDRNMGTKRAKCVDLAAQIKANIDANGVNCKSPDTATLLGIYSRQYRFSPLVDLIAATNFQRPKKQWLRLPLRLAKHDSAYEEGMYITVEECDDTAVA"
join(8915..9052,9216..9726,10663..10911,10976..11174,11235..11542,11596..>12210)
/gene="CG11866"
/product="Ct37014"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
<8915..>12210
/gene="CG11866"
/map="46E4-46E6"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
join(9286..9726,10663..10911,10976..11174,11235..11596..>12210)
/gene="CG11866"
/note="CG11866 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
/protein_id="AAF58839.1"
/db_xref="GI:7303791"
/translation="MVSTQMCSSGGTSASEGVSPFGAGSGSASCTGMASGAVSSGSGTGVQPEPSTANRSVLRRSTSYOGNOAHSSYRVVSAGSLAESASVTRLSHYTMQMQRPPTATTTTTSAPSLYDLNPLIEKILSYDYKKNVSNLRLVSHRNDICTMAMLNFTAKOIKTTLRSFOAIKASMPRRSHRNHPHCECDIETCYRLSLLOSMGKHIERHCFFFGAILDEVQAILNYSITPRQRYRYTDELSTFAMAYFKDRIEATLPGLAYKDFDTPTTKRTPLAISDLESDSNSSPPQNMMVLRKGIKRIQKMKMYNNOLSVLRTLELRTCRAEQSKOLAESNLISEQOKQTLVYANRLDENDKKEEMSRKSTLLQELNCKTELQFWRKSPALPVCSSCNOKVAVVPPEDQALVQGVDPHFIIINEDVAESDVSGELKEFASPESSTAKLAVSTAKNLRKMPSESQSNATATSKAAVDSQPMATGAAKAGGYSPLKFLYGNHQHSQVIVSPVSNLGVAAATDNYVSEALEEPEEKARRVQKANNRYNTPGKRSK"
complement(join(<12115..12937,12999..13642,13694..13708))
/gene="CG17753"
/product="Ct11457"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(<12115..>13708)
/gene="CG17753"
/map="46E4-46E6"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(join(12802..12937,12999..13642,13694..13708))

WLAQAOKANTDANGVNCKSPDTATLLGIYSRQYRFSPLVDLIAETNFDRIPKKQMWLRPLRLILAKHDSAYEEGMYITVEECDDTAVA"
join(2152..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8315)
/gene="Pfk"
/product="Ct13275"
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
join(2613..2759,3801..4142,4200..4354,4727..4835,6101..6289,6499..6941,6999..7841,7921..8059)
/gene="Pfk"
/note="Pfk gene product [alt 2]"
/codon_start=1
/db_xref="FLYBASE:FBan0004001"
/db_xref="FLYBASE:FBgn0003071"
/protein_id="AAF58841.1"
/db_xref="GI:7303793"
/translation="MNSEINQRELARGSQDKGLAVTSGDSQGMNAVRACVYRMAIYLGCXVYREGYQGVGGGAYLGTKRTLPKGKFEIARLKEFKIQGLITIGFESYHAAGQIADQRNYPQFCIPVIVIPSTISNNVPGTFEFLGCDTGLNEITEICDRIRQAQGTKRRVVFETWGGYCYLATLAGAGDAAYIYEKFSIKLOQDVYHMAKMAEVSRLILRKEKASENSTDFLYRLYSEEGGLTCRNNILGHMQQGSPTPDRNMGTKRAKCVDLAAQIKANIDANGVNCKSPDTATLLGIYSRQYRFSPLVDLIAATNFQRPKKQWLRLPLRLAKHDSAYEEGMYITVEECDDTAVA"
join(8915..9052,9216..9726,10663..10911,10976..11174,11235..11542,11596..>12210)
/gene="CG11866"
/product="Ct37014"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
<8915..>12210
/gene="CG11866"
/map="46E4-46E6"
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
join(9286..9726,10663..10911,10976..11174,11235..11596..>12210)
/gene="CG11866"
/note="CG11866 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0011866"
/db_xref="FLYBASE:FBgn0033486"
/protein_id="AAF58839.1"
/db_xref="GI:7303791"
/translation="MVSTQMCSSGGTSASEGVSPFGAGSGSASCTGMASGAVSSGSGTGVQPEPSTANRSVLRRSTSYOGNOAHSSYRVVSAGSLAESASVTRLSHYTMQMQRPPTATTTTTSAPSLYDLNPLIEKILSYDYKKNVSNLRLVSHRNDICTMAMLNFTAKOIKTTLRSFOAIKASMPRRSHRNHPHCECDIETCYRLSLLOSMGKHIERHCFFFGAILDEVQAILNYSITPRQRYRYTDELSTFAMAYFKDRIEATLPGLAYKDFDTPTTKRTPLAISDLESDSNSSPPQNMMVLRKGIKRIQKMKMYNNOLSVLRTLELRTCRAEQSKOLAESNLISEQOKQTLVYANRLDENDKKEEMSRKSTLLQELNCKTELQFWRKSPALPVCSSCNOKVAVVPPEDQALVQGVDPHFIIINEDVAESDVSGELKEFASPESSTAKLAVSTAKNLRKMPSESQSNATATSKAAVDSQPMATGAAKAGGYSPLKFLYGNHQHSQVIVSPVSNLGVAAATDNYVSEALEEPEEKARRVQKANNRYNTPGKRSK"
complement(join(<12115..12937,12999..13642,13694..13708))
/gene="CG17753"
/product="Ct11457"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(<12115..>13708)
/gene="CG17753"
/map="46E4-46E6"
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
complement(join(12802..12937,12999..13642,13694..13708))

mRNA

CDS

mRNA

gene

CDS

mRNA

gene

CDS

```
/gene="CG17753"
/note="CG17753 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0017753"
/db_xref="FLYBASE:FBgn0033487"
/protein_id="AAF58838.1"
/db_xref="GI:7303790"
/translation="MSSIKIEFAVMRRGDSYAGALRSALDGVGVQVEIDTQGRVVI
QTPRWEIQDKIEATGVRVLSFGGQSVALINTGSSVVDKTPIQGVVRFITAD
KPGVVDVGVVGLHGLHIESGDSVSGEHYNPRQSPHGPSPAAGAEERH
AGDLGNTRADENGRAITFVDPVLEVMVIGRAVVTANADDLGRGGNDQSLDNGS
ERIAACGIIARSAGILENFKRICACDGVTLWDERNKPLAGKERSOKL"
join(<18636..18761,19765..19876,19968..20213,22806..22878,
22958..23344,23489..23633,23934..24407,25026..>25202)
/gene="CG11867"
/product="CT37016"
/db_xref="FLYBASE:FBan0011867"
/db_xref="FLYBASE:FBgn0033488"
<18636..>25202
/gene="CG11867"
/map="46E4-46E8"
/db_xref="FLYBASE:FBan0011867"
/db_xref="FLYBASE:FBgn0033488"
join(18636..18761,19765..19876,19968..20213,22806..22878,
22958..23344,23489..23633,23934..24407,25026..>25202)

alignment_scores:
  Quality: 1136.00      Length: 862
  Ratio: 2.197          Gaps: 31
Percent Similarity: 59.97 Percent Identity: 34.33

alignment_block:
US-09-434-382-2 x AE003830/rev ..
Align seg 1/1 to reverse of: AE003830 from: 1 to: 261846

9 ArgSerAlaAlaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAl 25
|||||
200687 AGGTCGCCGATCTACAGAACCTTG...CGAACGCTCAACACCACTAATTT 200641

25 aProAlaArgGluArgProGlyAspProLeuArgHisLeuArgT 42
|||||
200640 GATGGCGGCCACAAATTCCTCGCCCAAGATCCTTTA.....ACAGGAC 200597

42 hrArgGluLysArgGlyProSerGly.....CysSerGly 53
|||||
200596 CTCGCTATGAGCGGGAACCAATGCTTAGAAAAAGCTGGCCTCGTG 200547

54 GlyProAsnThrValTyrLeuGlnValAlaAlaGlySerArgaspse 70
|||||
200546 GTGCCGGGAACAGTAATCTCAGGTTCTGGGCTCTGGCGCCATGGAGC 200497

70 rGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnC 87
|||||
200496 GCCTCGCGCTGTTATTGTTTACGACGACGACGACGCTATCTCTCAACT 200447

87 ysgGlyGluGlyValGlnArgLeuMetGlnGlnHisLysLeuLysValAla 103
|||||
200446 GCGCGGAGGAACCCAGAGACTAGCTACGAGCACAACCCGCCCTATCA 200397

104 ArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyG1 120
|||||
200396 CGCCTGGAACAAATTTTCCTACACAAAATACGTTGGGCTCTCTGCGGAGG 200347

120 yLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysV 137
|||||
200346 GTTACCAGGTTTGACGCTCACCATCAAGATCCGGAGTCCGGGACATTG 200297

137 alLeuSerGlyProProGlnLeuLysTyrLeuGluAlaIleLysIle 153
|||||
200296 GACTCCATGAGCCGCTCATCTCGGCTCTATGCTGCAGTCAATGCCGACGT 200247

154 PheSerGlyProLeuLysGlyIleGluLeuAlaValArgProHisSerAl 170
|||||
```

```
200246 TTCGTG...GTGCTAAAGATCTCCAG.....GTGCGGCCAAACGATTG 200206
|||
170 aProGlu.....TyrGluAspGluThrMetThrValTyrGlnIleP 184
|||
200205 TTCTGAGGGCCCTGTTTTCAGGATTCATCTTGAAGGTGGACTCTGCG 200156
|||
184 ro...IleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSer 199
|||
200155 CCCTGATCAATTCCGAA..... 200139
200 ProGluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGl 216
|||
200138 .....GATCCGACGAAAGT..... 200124
216 uSerAsnGluAsnGluProHisLeuProHisGlyValserGlnArgArg 233
200124 ..... 200124
233 lyValArgAspSerSerLeuValAlaPheIleCysLysLeuHisLeu 249
|||
200123 .....GTAATTAACTACATTTGTCAGCTGAAACCG 200094
250 LysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProVa 266
|||
200093 CGAGCTGGCGCCCTAAACTTGGTGAAGTGGCTGGAGCAAGGTGTGCGCC 200044
266 lGlyThrAlaAlaIleAlaProIleIleAlaValLysAspGlyLys 283
|||
200043 AGGA.....CCACTTCTGGGACACCTTAAGAAATGCCAACG 200009
283 erIleThr...HisGluGlyArgGluIleLeuAlaGluGluLeuCysThr 298
|||
200008 ATATCACACTGCCTGATGTAAGGTGCTGCTTCAGTCGACGTCACCGAA 199959
299 ProProAspProGlyAlaAlaPheValValGluCysProAspGluSe 315
|||
199958 GCCAGTGAACACTGCATCTCTTTGTTTCTGATGTCGCCAGTGAGAA 199909
315 rPheIleGlnProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyL 332
|||
199908 CTATCTACACGACTGCTTACACACGCAAGGCTTAAAGATTTAGCG 199859
332 ysAlaAsp...AlaProValAlaLeuValValHisMetAlaProAlaSer 347
|||
199858 AAGAGAAGTTAACTGAAGTGGCGCTTGTGTGCACTTCACGCCCTATCAT 199809
348 ValLeuValAspSerArgTyrGlnGlnTrp...MetGluArgPheGlyPr 363
|||
199808 ATATCTCTCTGCTCAGGTGTACAAGATTTTGTGTGGAGAGTTTTCATC 199759
363 oAspThrGlnHisLeuValLeuAsnGlu.....AsnCysAlaSerValH 378
|||
199758 TGAGGCACAGCACATATACTTAAGTTCTCCCTCAACCAATTTCCAGCT 199709
378 isAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLeuIleHisPro 394
|||
199708 AT...GCAGCGGCCCATCGCATTCACACCACTCCACCACTGGCACCG 199662
395 AspIlePheProLeuLeu...ThrSerPheArgCysLysLysGluGlyPr 410
|||
199661 CAAGTATTTCACCTCTCTGGCGGAACAATTTGCTCTGCCAGAGTCAG.... 199617
410 oThrLeuSerValProMetValGlnGlyGlu..... 420
|||
199616 .ACTCTCAGTCTAAACCTTTAAAAAACAAGATTGGACAGGCCGATTCGG 199568
420 ..... 420
199567 AGGACAAAGCAATTCGAAAGGCTAACGAGACAGACAGAACAGGGTGTGGTC 199518
421 CysLeuLeuLysTyrGlnLeuArgProArgGluTrpGlnArgAspAl 437
|||
```

199517 GCATGACAACTATACCTGAGACCAAGAAAGTTGGTACTATTATTTT 199468
437 aile...ileThrCys..... 441
199467 ACTTACCATTCATGCTATTAAGTATATTTTATTACCTGAAGCTCGGAT 199418
442Asn.ProGluGluPheIleValGluAlaLe 451
199417 CGCACTCTTTGAGTCCAACTACACAGAGGAGTATGTTAAGGAGACCCA 199368
451 uGlnLeuProAsnPhcGlnGlnSerValGlnGluTyrArgArgSerAlaG 468
199367 TCCGCTACAGAGATTCTGGAATATTGGCCAAATTCAAA.....G 199327
468 InAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleIle 484
199326 AGAGTATAGTTTCCCTGCAATTCGCGCATAGCTACCCGAAGAATAAT 199277
485 PheLeuGlyThrGlySerAlaIlePrometLysIleArgAsnValSerAl 501
199276 TTCTTGGGCACAGGCTCCGTCATCCCAATAAGACCGCAACCTGAGCTC 199227
501 aThrLeuValAsnIleSerProAspThrSerLeuLeuLeuLeuLeuLeu 518
199226 CATTCGTATCAGGACTGCAATTCGATGCTTACGCTTGTGTTGATTTGGAG 199177
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArg 534
199176 AAGGAACCTATGCTCAATTTGACGACTATATGGACACGAAAGGCGAG 199127
535 ...ValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAs 550
199126 TTGATTCITCGCCAGCTACAGCGGATTTATGATCCATTTGCATTCGCCA 199077
550 phisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuA 567
199076 CCACCACATTTGATTGATTGTTGCTGCTCGGAAAGGAGCAATTA.... 199031
567 laSerLeuGlyLysPro...LeuHisProLeuLeuValValAlaProAsn 582
199030AAACCTAGGCGAGACCCCATTTTCTGCTGGCTCCTCGC 198992
583 GlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLe 599
198991 CAAATTTGAACCTTGCTGGAGTTCTACATCGACAATAAGAACCGTACA 198942
599 uHisHisIleSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluI 616
198941 AGATCGGTATACCTAGTG.....GGCAACGGTGAACCTCC 198907
616 leSerSerPro.....AlaValGluArgLeuIleSerSerLeu 628
198906 TAGCTAGTCCCTGAGCGGTGAACAAAGTTGAACGCTA..... 198869
629 LeuArgThrCysAspLeuGluGluPheGlnThrCysLeuValArgHisCy 645
198868GGAATTACGTCCTATCCACCTGCCTAGTAGTACGCACTG 198831
645 SlyHisAlaPheGlyCysAlaLeuVal.....HisThrS 657
198830 CCCCACACTCTTCGGAATTAAGCCTAACTCTGGCGGCAAGACAAATAGCG 198781
657 erGlyTrpLysValValTyrSerGlyAspThrMetProCysGluAlaLeu 673
198780 AACCCGTCACATCATTACAGCGCGACACTATGCCATGTCAGGATCTG 198731
674 ValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuCl 690
198730 ATCGATCTGGCCCGCGACTCCACCGTCTTATTACAGAGCGCACAAATGGA 198681
690 uAspGlyLeuGluGluAlaValGluLysThrHisSerThrThrSerg 707
198680 GGATGACTTGGAGGAGGCGCGGTTAAAGACGACACAGCACCGCTGTGCGC 198631

707 InAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsn 723
198630 AGGCTATTACGAGCGCGGATATGAACCGCCCATAGATTCTCTCC 198581
724 HisPheSerGlnArgTyrAlaLysValPro...LeuPheSerProAsnPh 739
198580 CACTTTTCGAGCGTTACGCTAAGTCCCGGATTACCGAGTGATGAGGA 198531
739 eSerGluLysValGlyValAlaPheAspHisMetLysValCysPheGlyA 756
198530 TATGCAACGAGTTGCAATAGCCTTTGATATATGGAAGTGACCGTTGAGG 198481
756 spPheProThrMetProLysLeuIleProProLeuLysAlaLeuPheAla 772
198480 ATCTGCAGCACTACCAACAAGCTCTACCCCGCCCTTTTCGCGATGACGCC 198431
773 GlyAspIleGluGluMetGluArgArgGluLysArgGluLeuArgG1 789
198430 GAGTACACGAAGAACTGGAGCAGCGGCTGTCAAGCGGAGCTGAACA 198381
789 nValArgAlaAlaLeuLeuSerArgGluLeuAla 800
198380 GGAGCGA.....AAGCGAAACTGGCA 198359
seq_name: gb_htg6:AC017383

seq_documentation_block:
LOCUS AC017383 115873 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces
ACCESSION AC017383
VERSION AC017383.1 GI:6553603
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10210513 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
Location/Qualifiers
1..115873
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 31582 a 25315 c 25855 g 33121 t
ORIGIN

alignment_scores:
Quality: 1132.50 Length: 837
Ratio: 2.234 Caps: 30
Percent Similarity: 60.573 Percent Identity: 34.528
alignment_block:
US-09-434-382-2 x AC017383 ..

Align seg 1/1 to: AC017383 from: 1 to: 115873
34 LysAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSerg1 50
62335 AAGGATCCTTTA.....ACAGGACCTCGCTATGAGCGGGAACCAATGT 62378
50 Y.....CysSerGlyGlyProAsnThrValTyrLeuGlnV 62

608 LysCysLeuGlnGluGlyAlaGluIleSerSerPro.....Al 620
:||||| :|||||
64028GGCAACGGTAACTCTAGTAGTCCCTCGAGCGGTGAACA 64068

620 aValGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluGluP 637
||||| :|||
64069 AGTTGAACGTCTA.....GGAAATACCTCCA 64094

637 heGlnThrCysLeuValArgHisCysLysHisAlaPheGlyCysAlaLeu 653
:||||| :||||| :||||| :|||||
64095 TATCCACCTGCCTAGTAGTACGACCTGCCCACTCTTCGGAATAAGCCTA 64144

654 Val.....HisThrSerGlyTrpLysValValTyrSerG 665
:||||| :||||| :||||| :|||||
64145 ACTCTGGCGGCAAGACCAATACGCAACCGCTCAAGATCACTTACAGCGG 64194

665 YAspThrMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrL 682
||||| :||||| :||||| :||||| :|||||
64195 CGACACTATGCCATGTCAGAGTCTGATCGATCTGGCGCGGACCTCCACCG 64244

682 euLeuIleHisGluAlaThrLeuGluAspGlyLeuGluGluAlaVal 698
:||||| :||||| :||||| :||||| :|||||
64245 TTCTTATTCACGAGGACAAATGGAGGATGACTTGGAGGAGGCGCGG 64294

699 GluLysThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgMe 715
||||| :||||| :||||| :||||| :|||||
64295 TTAAGACGACGACGACCGCTGTCGAGGCTATTTCAGCAGGCGGAATAT 64344

715 tAsnAlaGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysV 732
||||| :||||| :||||| :||||| :|||||
64345 GAAGCGCGCCCATACGATTTCTCCACATTTTCGACAGCTTACGCTAAGT 64394

732 alPro....LeuPheSerProAsnPheserGluLysValGlyValAlaPhe 747
||| ||| ||| ||| :||| :||||| :||||| :|||||
64395 GCCCGGATTACCGATGATGAGGATATCGACGAGTTGCAATAGCTTT 64444

748 AspHisMetLysValCysPheGlyAspPheProThrMetProLysLeuI 764
||||| :||||| :||||| :||||| :|||||
64445 GATAATATGAAGTGAACCGTTCGAGGATCTGCAGCACTACCAAGCTCTA 64494

764 eProLeuLysAlaLeuPheAlaGlyAspIleGluGluMetGluGlu 781
||| ||| ||| :||||| :||||| :||||| :||||| :|||||
64495 CCGCGCCCTTTTCGGATGATGACCGGATACCGGANGAAGCTGGACAGC 64544

781 rGArgGluLysArgGluLeuArgGlnValArgAlaAlaLeuLeuSerArg 797
|| :||||| :||||| :||||| :|||||
64545 GGGCTGTCAAGCGGAGCTGAAACAGGACGA.....AAGCGA 64582

798 GluLeuAla 800
:|||||
64583 AAACGTGGCA 64591

seq_name: gb_hgt1.AC007417

seq_documentation_block:
LOCUS AC007417 125448 bp DNA HTG 02-AUG-1999
DEFINITION Drosophila melanogaster chromosome 2 clone BACR48F07 (D625) RPOI-98
48.F.7 map 47A-47B strain Y; cn bw sp, *** SEQUENCING IN PROGRESS
***, 94 unordered pieces.

AC007417
AC007417.3 GI:5670592
HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Rubin,G.M., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.

TITLE
JOURNAL
REFERENCE
AUTHORS

Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Rubin,G.M., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Sequencing of Drosophila melanogaster
unpublished
2 (bases 1 to 125448)
Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
Rubin,G.M., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.

TITLE
JOURNAL
COMMENT

Submitted (28-APR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Aug 2, 1999 this sequence version replaced gi:5629944.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 94 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 969: contig of 969 bp in length
* 970 1049: gap of unknown length
* 1050 2078: contig of 1029 bp in length
* 2079 2158: gap of unknown length
* 2159 3249: contig of 1091 bp in length
* 3250 3329: gap of unknown length
* 3330 4547: contig of 1218 bp in length
* 4548 4627: gap of unknown length
* 4628 5262: contig of 635 bp in length
* 5263 5342: gap of unknown length
* 5343 6096: contig of 754 bp in length
* 6097 6176: gap of unknown length
* 6177 7486: contig of 1310 bp in length
* 7487 7567: gap of unknown length
* 7568 8161: contig of 594 bp in length
* 8162 8241: gap of unknown length
* 8242 9090: contig of 850 bp in length
* 9091 9170: gap of unknown length
* 9171 9903: contig of 733 bp in length
* 9904 9983: gap of unknown length
* 9984 10564: contig of 581 bp in length
* 10565 10644: gap of unknown length
* 10645 11655: contig of 1011 bp in length
* 11656 12459: gap of unknown length
* 12460 12539: contig of 724 bp in length
* 12540 13197: gap of unknown length
* 13198 13278: contig of 658 bp in length
* 13279 14314: gap of unknown length
* 14315 14394: contig of 1037 bp in length
* 14395 15048: gap of unknown length
* 15049 15128: contig of 654 bp in length
* 15129 16091: gap of unknown length
* 16092 16171: contig of 963 bp in length
* 16172 16817: gap of unknown length
* 16818 16897: contig of 646 bp in length
* 16898 17598: gap of unknown length
* 17599 17678: contig of 701 bp in length
* 17680 18750: gap of unknown length
* 18751 18750: contig of 1072 bp in length

```
* 18751 18830: gap of unknown length
* 18831 19376: contig of 546 bp in length
* 19377 19456: gap of unknown length
* 19457 20255: contig of 799 bp in length
* 20256 20335: gap of unknown length
* 20336 20957: contig of 622 bp in length
* 20958 21037: gap of unknown length
* 21038 22187: contig of 1150 bp in length
* 22188 22267: gap of unknown length
* 22268 22871: contig of 604 bp in length
* 22872 22951: gap of unknown length
* 22952 24044: contig of 1093 bp in length
* 24045 24124: gap of unknown length
* 24125 25190: contig of 1065 bp in length
* 25191 25269: gap of unknown length
* 25270 26217: contig of 948 bp in length
* 26218 26297: gap of unknown length
* 26298 27312: contig of 1015 bp in length
* 27313 27392: gap of unknown length
* 27393 28934: contig of 1542 bp in length
* 28935 29014: gap of unknown length
* 29015 29998: contig of 984 bp in length
* 29999 30078: gap of unknown length
* 30079 31456: contig of 1378 bp in length
* 31457 31536: gap of unknown length
* 31537 32422: contig of 1706 bp in length
* 32423 33322: gap of unknown length
* 33323 34819: contig of 1497 bp in length
* 34820 34899: gap of unknown length
* 34900 35765: contig of 866 bp in length
* 35766 35845: gap of unknown length
* 35846 37032: contig of 1187 bp in length
* 37033 37112: gap of unknown length
* 37113 38879: contig of 1767 bp in length
* 38880 38959: gap of unknown length
* 38960 39889: contig of 930 bp in length
* 39890 39969: gap of unknown length
* 39970 40967: contig of 998 bp in length
* 40968 41047: gap of unknown length
* 41048 42656: contig of 1609 bp in length
* 42657 42736: gap of unknown length
* 42737 44172: contig of 1436 bp in length
* 44173 44252: gap of unknown length
* 44253 45669: contig of 1417 bp in length
* 45670 45749: gap of unknown length
* 45750 47740: contig of 1991 bp in length
* 47741 47820: gap of unknown length
* 47821 48827: contig of 1007 bp in length
* 48828 48907: gap of unknown length
* 48908 49770: contig of 863 bp in length
* 49771 49850: gap of unknown length
* 49851 51951: contig of 2101 bp in length
* 51952 52031: gap of unknown length
* 52032 53654: contig of 1623 bp in length
* 53655 53734: gap of unknown length
* 53735 55071: contig of 1337 bp in length
* 55072 55151: gap of unknown length
* 55152 56770: contig of 1619 bp in length
* 56771 56850: gap of unknown length
* 56851 58743: contig of 1893 bp in length
* 58744 58823: gap of unknown length
* 58824 60291: contig of 1468 bp in length
* 60292 60371: gap of unknown length
* 60372 63759: contig of 3388 bp in length
* 63760 63839: gap of unknown length
* 63840 66476: contig of 2637 bp in length
* 66477 66557: gap of unknown length
* 66558 69321: contig of 2765 bp in length
* 69322 69401: gap of unknown length
* 69402 70806: contig of 1405 bp in length
* 70807 70886: gap of unknown length
* 70887 73742: contig of 2856 bp in length
* 73743 73822: gap of unknown length

* 73823 75075: contig of 1253 bp in length
* 75076 75155: gap of unknown length
* 75156 77979: contig of 2824 bp in length
* 77980 78059: gap of unknown length
* 78060 81375: contig of 3316 bp in length
* 81376 81455: gap of unknown length
* 81456 84452: contig of 2997 bp in length
* 84453 84532: gap of unknown length
* 84533 87596: contig of 3064 bp in length
* 87597 87676: gap of unknown length
* 87677 102529: contig of 14853 bp in length
* 102530 102609: gap of unknown length
* 102610 103128: contig of 519 bp in length
* 103129 103208: gap of unknown length
* 103209 103812: contig of 604 bp in length
* 103813 103892: gap of unknown length
* 103893 104112: contig of 220 bp in length
* 104113 104192: gap of unknown length
* 104193 104767: contig of 575 bp in length
* 104768 104847: gap of unknown length
* 104848 105570: contig of 723 bp in length
* 105571 105650: gap of unknown length
* 105651 106151: contig of 501 bp in length
* 106152 106231: gap of unknown length
* 106232 106941: contig of 710 bp in length
* 106942 107021: gap of unknown length
* 107022 107705: contig of 684 bp in length
* 107706 107785: gap of unknown length
* 107786 108412: contig of 627 bp in length
* 108413 108492: gap of unknown length
* 108493 109249: contig of 757 bp in length
* 109250 109329: gap of unknown length
* 109330 110040: contig of 711 bp in length
* 110041 110120: gap of unknown length
* 110121 110727: contig of 607 bp in length
* 110728 110807: gap of unknown length
* 110808 111413: contig of 606 bp in length

alignment_scores:
    Quality: 1128.00      Length: 811
    Ratio: 2.274         Gaps: 28
    Percent Similarity: 61.159      Percent Identity: 34.649

alignment_block:
US-09-434-382-2 x AC007417/rev ..

Align seg 1/1 to reverse of: AC007417 from: 1 to: 125448

55 ProAsnThrValTyrLeuGlnValValAlaAlaGlySerArgAspSergl 71
|||:::||||| |||||:::|||||:::|||||:::|||||:::
84410 CCGGGAACAGTAATCTTCAGGTTCTGGGCTCTGGCGCAATGGAGCGCC 84361

71 yAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysG 88
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
84360 TGGCGCTGTTTATTGTTTACGGACCGACCGCTATCTCTCAACTGCG 84311

88 lyGluGlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArg 104
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
84310 GCAGGGAACCCAGAGACTAGCTCAGCAGGACACAAACCCGCTATCAGC 84261

105 LeuAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLe 121
|||:::|||||:::|||||:::|||||:::|||||:::|||||
84260 CTGGAACAAATTTTCCTTACACAAATACGTGGGCTTCCTCGCGAGGTT 84211

121 uSerGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysVal 138
| |||||:::|||||:::|||||:::|||||:::|||||:::
84210 ACCAGTTTGACGCTCACCATCCAAAGATCGCGAGTGGCGGACATTCGAC 84161

138 euSerGlyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePhe 154
|| r |||||:::|||||:::|||||:::|||||:::|||||
84160 TCATGAGACCGCTCACTCTCGGCTCTATGCTGCAGTCAATGCGAGTTTC 84111

155 SerGlyProLeuLysGlyIleGluLeuAlaValArgProHisSerAlaPr 171
```



```
repeat_region /rpt_family="LTR16C" 12453..12725
repeat_region /rpt_family="AluJb" 13377..13785
repeat_region /rpt_family="MSTB" 14118..14153
repeat_region /rpt_family="(CA)n" complement(15320..15620)
repeat_region /rpt_family="AluJo" 15683..15759
repeat_region /rpt_family="(CA)n" complement(15838..15893)
repeat_region /rpt_family="MER91A" complement(15915..16080)
repeat_region /rpt_family="MER91A" complement(16474..16544)
repeat_region /rpt_family="L2" 17907..17920
repeat_region /rpt_family="L1ME" 18271..18311
repeat_region complement(18368..18529)
repeat_region /rpt_family="AluJb" complement(18530..18560)
repeat_region /rpt_family="(CA)n" complement(18561..18681)
repeat_region /rpt_family="AluJb" 18760..18965
repeat_region /rpt_family="AluSg" complement(19638..19896)
repeat_region /rpt_family="MIR" complement(19971..20260)
repeat_region /rpt_family="L1ME3A" 20263..20552
repeat_region /rpt_family="MER58B" complement(20553..20816)
repeat_region /rpt_family="L1ME3A" complement(20821..21069)
repeat_region /rpt_family="L1M4" complement(21071..21100)
repeat_region /rpt_family="(TAAA)n" complement(21101..21338)
repeat_region /rpt_family="AluSx" complement(21348..21598)
repeat_region /rpt_family="L1M4" 21799..22104
repeat_region /rpt_family="AluJo" 22151..22562
repeat_region /rpt_family="MSTA" 23978..24287
repeat_region /rpt_family="AluSg" 24852..25199
repeat_region /rpt_family="L2" complement(25697..25985)
unsure /rpt_family="MLR1D" 25816..25895
repeat_region /note="Single-stranded coverage." complement(25994..26099)
repeat_region /rpt_family="purine-rich" complement(26122..26291)
repeat_region /rpt_family="MLR1D" 26312..26337
repeat_region /rpt_family="(CAAA)n" 26456..26480
repeat_region /rpt_family="(GGGA)n" complement(26576..26781)
repeat_region /rpt_family="MER6B" complement(26912..26990)
repeat_region /rpt_family="MER21B" 26991..27290
repeat_region /rpt_family="AluSx" complement(27291..27975)
repeat_region /rpt_family="MER21B" 28034..28197
repeat_region /rpt_family="LTR28" 28266..28795
repeat_region /rpt_family="LTR28" 28843..28916
repeat_region /rpt_family="LTR28" 28920..29005
repeat_region /rpt_family="MIR" 29140..29233

alignment_scores:
  Quality: 1012.00      Length: 1315
  Ratio: 3.057          Gaps: 14
  Percent Similarity: 25.171  Percent Identity: 24.487

alignment_block:
  US-09-434-382-2 x AC005277

Align seg 1/1 to: AC005277 from: 1 to: 118788
484  ilePheLeuGlyThrGlySerAlaileProMetLysileArgAsnValse 500
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99759 ATCAGCCTCTGAACCATCAGCAGCTCTTCCTAGT.....GGCAGTGACTC 99802
500 rAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysG 517
    | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99803 TCTTCCTCTCTCTCTGCAGCCCGACACAGCTCTGCTACTGGACTGTG 99852
517 lyGluGlyThrPheGlyGlnLeuCysArgHisTyrglyAspGlnValasp 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99853 GTGAGGCGACATTTGGGAGCTGTGCCGTCATTTACGGAGACCAGGTGGAC 99902
534 ArgValLeuGlyThrLeuAlaValPheValSerHisLeuHisAlaAs 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99903 AGGGCTCTGGGCACCTGGCTGCTGTGTTGTGTCCTCCACCTGCACGCAGA 99952
550 pHisHisThr..... 553
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
99953 TCACCACAGGTGAGTGTGGGCTGGACCAACAAAGCTGGAGCTGGAGGA 100002
553 ..... 553
100003 GGCAGTGCACAGTTGAGTTGGCCCTTTGGCTGGCTGGCTTTTCCTCCGCTTC 100052
553 ..... 553
100053 CAAACTTGCCAGAGCTTTTGTACTCATCTCTGGCTAGAGAAATGCTTTT 100102
553 ..... 553
100103 TTGCAAAACTCAACATAGTCCTTCTGCGCCACAAGAATGCTTCTCTTCC 100152
553 ..... 553
100153 TCTTCAGTTCTTCTCTGACGAGGACAGGTTTGAGTTTACCCAGCCTTC 100202
553 ..... 553
100203 CTTGAGCTTGAATCTCACACGCCCTGCTCAGCGGAAGCTTTGACCGGAT 100252
553 ..... 553
100253 GCAGGAGGTGTGGCTATGAGACCCTCACCTTGGTCTCTCTGGGTGCGGG 100302
553 ..... 553
100303 CCCTGGGCGGTTGCCCTCTTCCAGCAGGGTCTGTCGCTTTTCTGCTCTG 100352
553 ..... 553
100353 TGACATTTACAGGCCATGGCGCAGGGGCTCGGCTCTGCGCACCACCCCACT 100402
553 ..... 553
```

100403 GCGGCTGTGTAGAGGCTGTGGGTGACGTGCGGCTGGCAACTCCTGTCAA 100452
553 553
100453 GAGAGAGGCTGCAGACCCCTAACCCGGAGGGATGGCCCTGGGGCCTGGC 100502
554 Gly. LeuProSerIleLeuLeuG 561
100503 TGACGCATGTCTCTGTCTTCTTGCACAGGCTTGCCAAAGTATCTGTGCTGC 100552
561 InArgGluArgAlaLeu..... 566
100553 AGAGAGAACCGCCTTGGTAAGTGTGGCAGCTTGATGGCGCTTCTGAGTTT 100602
566 566
100603 CAGCGGTTTACACATCATCCGCCATGCCCTTGGCACTCCAGTTTATT 100652
567 Ala. Ser 568
100653 GAGATGTTCTGTCTGAGTCGGCACTTGCATTTTGTTCAGGCATCT 100702
569 LeuGlyLysProLeuHisProLeuValAlaProAsnGlnLeuLy 585
100703 TTGGAAAGCCGCTTCACCCCTTGTGCTGGTGGTGGCCCAACCAAGCTCAA 100752
585 sAlaTrpLeuGlnGlnTrpHisAsnGlnCysGlnGluValLeuHisHis. 601
100753 AGCTGGCTCCAGCAGTACCACCAACCAAGTGCAGAGGCTCTGCACCACA 100802
601 601
100803 TCAGGTGAGCATCCAGGAGCCTGGCCCGCTGGCTGTGTTGCTGTGCC 100852
601 601
100853 GTCTCCTTCAGAGCTCAAGGTGGACACTGGGTAGTTACCAATATCCCC 100902
601 601
100903 CAGCAGCCTTGCCCTTGACATGGTCCAGATGGCAGAGCAGGGGAGAAG 100952
601 601
100953 TGCATTGGCTGAAGGACAGACAAACCATTAGATAGTTCCCATGTAATGCTTA 101002
601 601
101003 TTTTCTTAGAAGCATTTCTCCAGTCCCTCATTTGAGTCTTGAGCTGCTT 101052
601 601
101053 TCTAACTTCGACAGCTTTTCTTGATGACAGACTTCCAGAGCCACAGCAC 101102
601 601
101103 CCAAAATAGTGGCTAGCACAGAGAATGTCCATAGCAGGTGTGGCTAGCT 101152
601 601
101153 GGCAGGTGGACCATCCTCACCCCAAGGGAAGAGTCCCTCTGTGTA 101202
601 601
101203 GCCATCGGTGGCGGTGCTGCCTGAGCGGAGGAGCAGCATTCACCTGTCTGG 101252
601 601
101253 GTTTCCTCCAGTGCCTAGAGGCTTTGGTTGGCTCTTTATATTGACTG 101302
601 601

101303 CTGTTTCCTCATCATAGTACTATGATTAACTTAACTCATGTTTCTCCTAAGA 101352
601 601
101353 ATGATTTTGGGTTCCTCCAGCCAAAGACTTAAACTTTGGTTCCAGATGTC 101402
602 Ile. SerMetI 605
101403 CAAGAAACGTTTATTATCATTTTAAATGTTTGTCTTTTACAGTATGAT 101452
605 eProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerProAlaValG 622
101453 TCCTGCCAAATGCCCTCAGGAAGGGCTGAGATCTCCAGCTCTGCAGTGG 101502
622 LuArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluGlu..... 636
101503 AAAGATTGATCAGTTCCTGCTTGGCAACATGTGATTGGAGAGGTAAGG 101552
636 636
101553 GGCACAGCCGAGGCATCATGGGGCGAGGTGGGAGCAGAGCTGCAGAG 101602
637 PheGlnThrCysLeuValArgHisCysL 646
101603 CCTCCAGCCCCACCTTTCAGTTTCAGACCTGTCTGGTGGCGCAGTGA 101652
646 YsHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLysValVal 662
101653 AGCATGGTTTGGCTGTGCCCTGTGTGCACACCTCTGGCTGGAAAGTGGTC 101702
663 TyrSerGlyAspThrMetProCysGluAlaLeuValArg..... 675
101703 TATTCGGGGACACCATGCTGCGAGGCTCTGTCGGATGGGTGAGTA 101752
675 675
101753 GAGGAAGAAGCAAGCCACCTGAGTTGCTCTGGGTTTGTGTAGCTGGA 101802
675 675
101803 GGTCAATGCAGTGGGCTTGCAGGGAACGTCACGAGCAGGAGGAGACTC 101852
675 675
101853 AGTCCCCACCTCAGAGTCTCTGTTGTCTATCTAGTAGGACAGCCAG 101902
675 675
101903 GGCAGGGAGCTGAGTTGAGACCAGAGAAACAGCAGTGAAGGCC 101952
676 Met 676
101953 TGTGTGCCGCTCTCGCAGAGAACTCTGCCCTGATCCTTGTGCTGCTCTC 102002
677 . GlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly 692
102003 CAGGAAAGATGCCACCTCTCTGATACATGAAGCACCCCTCGAAGATGCT 102052
693 LeuGluGluGluAlaValGluLysThrHis..... 702
102053 TTGAAGAGGAAGCAGTGGAAAGACACACAGGTAGCAAGGCCGGTCTAG 102102
702 702
102103 TCCTTGTGCCCCACATCCTCTCCCTCCCCACTACGTGACACTGAGCAGC 102152
702 702
102153 CGTGGTTGTCTCCACTGATGTGGGGCTGCCCTGCTTCTTATCAAGGCT 102202
702 702
102203 ATGGGGCTTCTTGACCTGTGGCAGTGCTCACAGGCTCTTGGCCTTTAT 102252

supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.

CDS are numbered using the following system eg SPAC5H10.01c. SP (S. pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), C (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.

The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.

Cosmid c1d4 is overlapped at the 5' end by cosmid c1f3 and at the 3' end by cosmid c22f3.

FEATURES

source

```
1. 31201
   Location/Qualifiers
   1. 31201
   /organism="Schizosaccharomyces pombe"
   /strain="972h-"
   /db_xref="taxon:4896"
   /chromosome="I"
   /clone="cosmid c1d4"
   /map="IL"
```

misc_feature

```
1. 133
   /note="the sequence of this region is derived from cosmid c 1f3, the true end of c1d4 is at position 134, while the true end of c1f3 is at position 4137 in this sequence"
```

gene

```
7. 864
   /gene="SPAC1f3.11"
   /note="SPAC1D4.01"
```

CDS

```
7. 864
   /gene="SPAC1f3.11"
   /note="SPAC1D4.01, len:285"
   /codon_start=1
   /product="hypothetical 33.0 kd protein"
   /protein_id="CAB56125.1"
   /db_xref="GI:5912524"
```

```
/translation="MRQIFGCTGYMGNGYFQEGLEIPPTTKSIRMNSIHKKKSNRS
FRKRVFNEKEFLEDDNDNDILRQALEATKRRKIRNSIIGINAELLLNQETKKEK
QNTANPEHNEADOTSQSSKLEIAOLPTVEDRAKOTNEVDINTHLLNFVEKIKQE
RLAONYSENGENTNALNTKNESTVQNIKNLSHPNESHSTIRDAALGAIREVDLGIISTD
VDNLKRRKKRKKRKEKLDKRALRTSEDAARDEFIEKRLKPLSQDEESKGIYRRF
RVYKDGDTQD"
complement(join(1088..1636,1686..1873,1970..2090,
2137..2316))
/gene="SPAC1D4.02c"
complement(join(1088..1636,1686..1873,1970..2090,
2137..2316))
```

```
/gene="SPAC1D4.02c"
/note="SPAC1D4.02c, len:345, SIMILARITY:Rattus norvegicus,
Q35254, golgi peripheral membrane protein p65., (451 aa),
fasta scores: Opt: 493, E():4.7e-20, (34.4% identity in
305 aa)"
/codon_start=1
/label="SPAC1D4.02c"
/product="putative golgi peripheral membrane protein"
```

```
/protein_id="CAB56126.1"
/db_xref="GI:5912525"
/translation="MFGGLKNFIEKSEALAGIHRESDSCGFRVLKVENDSKAYNAR
IESYDFITAVNGILLDGPDMFALLRDSSPEVTFVLSLKQITRKVNKIKNSDEK
IGMVLQWASTAPADAIWHILNVDDSPVARSLVPYEDYIVGTPEGMTGKALSIDL
IESHLNPLRLIYINHYRDSRQVTIVPNRHGNGAIGCGVGHVLRHLPAPLSGPP
POPGDIVFSPMLGGPDHKVQSPETENFLTPRPKKTASANAGSSNIEISIPYORHK
KSHKGAIDSSIQSYLDEEKLRSRELDHKTADSNDSTQITPLPPPPVAVNSTNDE
SAPONEELVKN"
complement(1637..1651)
/note="splice branch and acceptor sequence,
ctaacaattatacag"
complement(1680..1685)
/note="splice donor sequence, gtagtgt"
complement(1874..1886)
/note="splice branch and acceptor sequence, ctaacgactacag"
```

```
/misc_feature
complement(1637..1651)
/note="splice branch and acceptor sequence,
ctaacaattatacag"
complement(1680..1685)
/note="splice donor sequence, gtagtgt"
complement(1874..1886)
/note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```
complement(1637..1651)
/note="splice branch and acceptor sequence,
ctaacaattatacag"
complement(1680..1685)
/note="splice donor sequence, gtagtgt"
complement(1874..1886)
/note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```
complement(1637..1651)
/note="splice branch and acceptor sequence,
ctaacaattatacag"
complement(1680..1685)
/note="splice donor sequence, gtagtgt"
complement(1874..1886)
/note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```
complement(1637..1651)
/note="splice branch and acceptor sequence,
ctaacaattatacag"
complement(1680..1685)
/note="splice donor sequence, gtagtgt"
complement(1874..1886)
/note="splice branch and acceptor sequence, ctaacgactacag"
```

misc_feature

```
/note="splice donor sequence, gtaagt"
```

misc_feature

```
complement(2091..2108)
```

misc_feature

```
/note="ctaagtgattttcacag, splice branch and acceptor"
```

gene

```
complement(2131..2136)
```

CDS

```
/note="gtagta splice donor sequence"
```

CDS

```
/gene="SPAC1D4.03c"
```

CDS

```
complement(join(2574..2820,2907..4201))
```

CDS

```
/gene="SPAC1D4.03c"
```

CDS

```
/note="SPAC1D4.03c, len:513, SIMILARITY:Fugu rubripes,
```

CDS

```
Q9YGN1 sand protein., (520 aa), fasta scores: opt: 609,
```

CDS

```
E():7.2e-32, (28.5% identity in 407 aa)"
```

CDS

```
/codon_start=1
```

CDS

```
/label="SPAC1D4.03c"
```

CDS

```
/product="hypothetical protein"
```

CDS

```
/protein_id="CAA93212.1"
```

CDS

```
/db_xref="GI:1177336"
```

CDS

```
/db_xref="SWISS-PROT:Q10150"
```

CDS

```
/translation="MEPTSEHSISKEEVENDVHRHSESGCSLLNPGNVLMAPSV
```

CDS

```
SEDDQEVSRSTPELRSHVENVEQLSDIHDSNPLNTSVSSSSNTAVDEIIKLL
```

CDS

```
SLISFDLAKQRKTYLIFSSGKPVFSNIVDDSDIEPTVGAQAIISSFEVSKKEETSF
```

CDS

```
STFSNVIVLKSPLLYLVGSPSTTLSAAYLLSELNLLYCOILGTAKAMQLTNSR
```

CDS

```
PNFDLRLLIGSNEQFLKELCQDNDYELVPLTNAISPLRASSFPDQSQLLRLTPK
```

CDS

```
SLLEFTAIRGLRCVMKAKKLLHANDYLLFLSFRQSFNDSMEHWVPVCFPTLN
```

CDS

```
PDAYIYISYFLCKDVTLMGSSGVFQSVKCAQVQAEIQDHQWKLKLYCEMD
```

CDS

```
RTTPRNPSCISHLYFSKYSQFTPGYSFSTPNFNTRTLYAIYASHDOAFHKKN
```

CDS

```
SFSINMTVHESLLFTWTSTAFDFHCIANATSSOLLIANVKNILRWIRRENRLFQ
```

CDS

```
TNLSF"
```

CDS

```
complement(2821..2836)
```

CDS

```
/note="splice branch and acceptor sequence,
```

CDS

```
ctaactatgtttta g"
```

CDS

```
complement(2901..2906)
```

CDS

```
/note="splice donor sequence, gtagtgt"
```

CDS

```
join(5400..5402,5439..5570,5618..7066)
```

CDS

```
/gene="SPAC1D4.04"
```

CDS

```
/note="cct2"
```

CDS

```
Join(5400..5402,5439..5570,5618..7066)
```

CDS

```
/gene="SPAC1D4.04"
```

CDS

```
/note="SPAC1D4.04, len:527"
```

CDS

```
/codon_start=1
```

CDS

```
/product="probable t-complex protein 1, beta subunit"
```

CDS

```
/protein_id="CAA93213.1"
```

CDS

```
/db_xref="GI:1177337"
```

CDS

```
/db_xref="SWISS-PROT:Q10147"
```

CDS

```
/translation="MSLNPHOIFNEQIERGENARLSFVGAIAGVGLVKTLGPKG
```

CDS

```
MDKLIQSNSSGDIYVNDGATILKSIALDNRKAAKVLNISKVQDDDEVGDGTSVCVFA
```

CDS

```
AEILLQAEIMVNAKIHPOVIIDGRIATKTAIDALRSSIDNSSDPKAFRSDLENIAR
```

CDS

```
TIGVNCPRKVMENALIANAMDTKVFGARVVDVTGKLAELEREREMKAKVE
```

CDS

```
KIKSHINCFINRQLIYNPQOLFADAGTMSIEHADFDGIERLSVLTGVEIASTFDHP
```

CDS

```
ELVKHGCKKIEEIIIGEDKMTKFSGVGAEACTIVLRGATHOLLDESERATHDALAV
```

CDS

```
LSQTVAESRVTLGGCAEMLKAKVEEATPEPKKAVAVSAFAKALSQLPILADNA
```

CDS

```
GFDSSELVAQLKAAHYDNGTMDGLDMEDEADMRKAKILEALKUKQAVSSGSEGAQ
```

CDS

```
LLLRVDITLKAAPRRER"
```

CDS

```
5425..5438
```

CDS

```
/note="splice branch and acceptor sequence,
```

CDS

```
tactaacgatatag"
```

CDS

```
Join(5529..5570,5618..7045)
```

CDS

```
/gene="SPAC1D4.04"
```

CDS

```
/note="Match to PF00118 cpn60_TCP1, TCP-1/cpn60 chaperonin
```

CDS

```
family Score 590.77"
```

CDS

```
5571..5576
```

CDS

```
/note="splice donor sequence, gtaagag"
```

CDS

```
5604..5617
```

CDS

```
/note="splice branch and acceptor sequence,
```

CDS

```
ttaacgattatag"
```

CDS

```
5741..5767
```

CDS

```
/gene="SPAC1D4.04"
```

CDS

```
/note="PS00995 Chaperonins TCP-1 signature 3"
```

CDS

```
complement(join(7282..8366,8419..8497))
```

CDS

```
/gene="SPAC1D4.05c"
```

CDS

```
complement(join(7282..8366,8419..8497))
```

```

/gene="SPAC1D4.05c"
/notes="SPAC1D4.05c, len:387, LOW SIMILARITY:Mus musculus,
Q920U0, polytropic murine leukemia virus receptor sygl.,
(695 aa), fasta scores: opt: 261, E():1.5e-10, (25.1%
identity in 398 aa) also shows low similarity to,
Saccharomyces cerevisiae, ERD1_YEAST, ERD1 protein., (362
aa), fasta scores: opt: 187, E(): 1e-05, (23.6% identity
in 382 aa)
SPAC1D4.05c, len:387, SIMILARITY:Mus musculus, Q920U0,
polytropic murine leukemia virus receptor sygl., (695 aa),
fasta scores: opt: 261, E():1.5e-10, (25.1% identity in
398 aa)"
/codon_start=1
/label=SPAC1D4.05c
/product="hypothetical protein"
/protein_id="CAA93214.1"
/db_xref="GI:1177338"
/db_xref="SWISS-PROT:Q10151"
/translation="MDLEVVEPFLHKKLALPFRIGLLVIVGTWLSVCYHLIYLNRY
QPISPRGSLSNKKWHLQIPLSNRHTDLEENTERKANLVSVPDFHAGYCFAAIISI
SWATGFILEKKTGGDIGLYSHPIYPLWITAFILVFPFMRYSRQGLKRSII
RVFLQADPRSPYKDFVSEIFTYAKALGDFYIFCVLQGHISKFTLPDLKCDGT
FVPLAMAYFPFVAILQCLHYGLSRKHKTEKINLLSALKHATAPVLYLSAIIHAKOT
KFTLTSGHGLFWLWILSALLSSAYTFLDWDFIDWRIRPFPHKSINHKKRPMEIYALG
CFINFLIRVWSMKLHPRLHQPHEYEMGFSEFMELIRFLWLFHLDIAISS"
/misc_feature
complement(8367..8389)
/notes="splice branch and acceptor sequence,
tactaacaactctcta tcttttag"

```

```

alignment_scores:
  Quality: 785.50      Length: 852
  Ratio: 1.650        Gaps: 32
  Percent similarity: 55.869  Percent identity: 28.404

alignment_block:
US-09-434-382-2 x SPAC1D4

Align seg 1/1 to: SPAC1D4 from: 1 to: 31201

59 TyrLeuGlnValAlaAlaGlySerArgAspSerGlyAla.....Al 73
||||| |||:||||:||||:||||:||||:||||:||||:||||:
14502 TATTTACAATTGTGAGCGTCAGTCTCGGGATACATCTGTATTCCTG 14551

73 aleuTyrValpheSerGluPheAsnArgTyrLeuPhe...AsnCysGlyG 89
||||| |||:||||:||||:||||:||||:||||:||||:||||:
14552 CATCCACTATTTTGTGATCTTAAGAGGTATGTATTTGGAAGCGTAGGG 14601

89 luGlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLeu 105
||||| |||:||||:||||:||||:||||:||||:||||:||||:
14602 AAGGTTGTCAAGAGCTATCTCTCCACAGCTTCGATTATCAAGATT 14651

106 AspAsnIlePheLeuThrArg..... 112
||||| |||:||||:||||:||||:||||:||||:||||:||||:
14652 AAAGATGCTCTTCTTAATCAAGAGCTTCATATCATCACCATACATA 14701

112 ..... 112
14702 TGATTTCTTCTTTCATCGTCTACTACTCTGTCTGTCTGATATGCTACAGC 14751

112 ..... 112
14752 TGGATGATCGGACAAAGTTATTGTATCTGAACGTAATAGTATGTGCTCG 14801

113 .....MethiTripsSerAsnValGlyGlyLeuSerGly.. 123
||||| |||:||||:||||:||||:||||:||||:||||:||||:
14802 ACTGTAACACTACTCTACATGCTGGGATCTTGTGCTGGGTTCCAGGTAA 14851

124 .....Metile.LeuThrLeu..... 128
||||| |||:||||:||||:||||:||||:||||:||||:||||:
14852 GTGTTAAATATATTAAATAATTCTGACAAATAGGCTTTTGTGCTCTC 14901

129 .....LysGluThrGly...LeuProLysCysValIle 138
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15752 ATTCACACACTTAAATTGTTAGATTCAAAAGTTTTTCTCTTTA.....GG

```

```

14902 AATGATATTTCGGAACCTCGGGAAACCGGAGAGCTAGTCCCTTTTGTCT 14951
138 uSerGlyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheS 155
||||| |||:||||:||||:||||:||||:||||:||||:||||:
14952 TCATGACCTAGTGAAGTTCATCAGTTTTTATCTTCGATCGCGCATTTCA 15001
155 erGlyProLeuLysGlyIleGluLeuAlaValArgProHisSer..... 169
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15002 CTTATCAT...ACGAATGTAATTTAAACGGTACAGGATATACTTCGGCT 15048
170 ...AlaProGluTyrGluAspGluThrMetThrValTyrGlnIleProI 185
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15049 GAGCCCCCAGTATTCGTTGATGAAATATTTGTAACTCCGCTAGTTGT 15098
185 e.....HisSerGluGlnArgA 191
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15099 TTCACCTAGTTAAATTTCTTTTAAAAAAGGAGCATGAAAAACATAACA 15148
191 rgGlyLysHisGlnProTrpGlnSerProGluArgProLeuSerArgLeu 207
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15149 GAGGTACAAAT.....GCAAGACCATTAAGAG... 15177
208 SerProGluArgSerSerAspSerGluSerAsnGluAsnGluProHisLe 224
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15178 .....GATAGAGCAATACTTCT.....CCTCATTTG 15203
224 uProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuValV 241
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15204 GTATTTCTCAGTTTCC.....ATGATACATCATTCGTCG 15238
241 al..... 241
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15239 TTGAGATGCAATGTATAACACTCGCGCTCTTTGGAAACGACAGCAAGCA 15288
242 .....AlaPheIleCysLysLeuHisLeuLysArgGlyAsnPh 254
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15289 GAGCTGTTTATTTCTTATTTATTTTCAATCCCTACCCAGGAAAT 15338
254 eLeuValLeuLysAlaLysGluMetGlyLeuProValGlyThrAlaAlaI 271
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15339 TCACGCCCAAGCAAAATCGCTTGGTATCACGAAAGGATTAGACTGT. 15387
271 leAlaProIleleAlaValLysAspGlyLysSerIleThrHisGlu 287
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15388 .....GGTAGACTCGCTCGAGGAGAACCTGTAACCTCTTGA 15423
288 ...GlyArgGluIleLeuAlaGluGluLeuCysThrProAspProGl 303
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15424 AACGAAACAGTATCTCCAAAAGAGTGATTTGCTCTTCTTATTTCTGG 15473
303 yAlaAlaPheValValValGluCysProAspGluSerPheIleGlnProI 320
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15474 CTCCTCTTTTATTTATTTATTTATTTGCCCCAACGAACTCGTAATTG 15523
320 leCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaPro 336
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15524 TCATTGAAATCATAAATGGACA.....AATGCACCTAAACCT 15561
337 ValAlaLeuValValHisMetAlaProAlaSerValLeuValAspSerAr 353
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15562 GTTGT...GTCAITTCATTCGTTACACCAAGATATACAAAATCCACG 15608
353 gTyrGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuVal 370
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15609 TTATCAGAGCTGGATTCTTCTTATTCCTCCCAAGTGAGGTGAGCATCTA 15658
370 euAsnGluAsnCysAlaSerValHisAsnLeu...ArgSerHisLysIle 385
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15659 CTTGACTGAAGTAAACGAAGTAAATAATATCTTAGAGTGCAGCTGCT 15708
386 GlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSe 402
||||| |||:||||:||||:||||:||||:||||:||||:||||:
15709 ATTCACACACTTAAATTGTTAGATTCAAAAGTTTTTCTCTTTA.....GG 15752

```

```

402 rPheArgCysLysLysGluGlyProThrLeuSerValProMetValGlnG 419
:|||||
15753 CTTAACTGC.....TATGAAGTTAAATAATGACAAA 15784

419 LyGlu.....CysLeuLeuLysTyrGlnLeuArg.....Pro 429
:|||||
15785 AAACAACGAAATTCGTTTGGCAACCCAAACCGCTTTCGGTTTGA 15834

430 ArgArgGluTrpGlnArgAspAlaIleLeuThrCysAsnProGluGluPh 446
:|||||
15835 AAAAAGACTGGTATTGACGATTGCGAAGTTGGGTTTCGATTGAAGATT 15884

446 eIleValGluAlaLeu...GlnLeuProAsnPheGlnGlnSerValGlnG 462
:|||||
15885 GAAGGATAAAATTTTGAAGAGAAACCAAGATTATAAATCATTTGTTGAAG 15934

462 lUtyrArgArgSerAlaGlnAspGlyPro...AlaProAlaGluLysArg 477
:|||||
15935 AGCTCAAAAATATGTTAGTAGATAAACCAAGCCAGCTTCAGGT 15984

478 SerGlnTyrProGluIleLeuPheLeuGlyThrGlySerAlaIleProMe 494
:|||||
15985 TCTGACATC...CAAATTGTACTCTTGCACCTGGATCGCAATGCCCTTC 16031

494 tIysIleArgAsnValSerAlaThrLeuValAsnIleSerProAsp.... 509
:|||||
16032 TTTATACAGAAATGTTTCATCTACTTACGTTAGAAATTCCTCGTTGACAAA 16081

510 .....ThrSerLeuLeuLeuAspCys 516
:|||||
16082 AATGTATGAAGACATCAGCTATCATGTATGAAACATACATTTTGGACTGT 16131

517 GlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyValAspGlnValAs 533
:|||||
16132 CGAGAGGCGACTTTGGGAGCTCTATCGAGGAGTATGGTGAATTTAA 16181

533 pArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAla 550
:|||||
16182 APACGAAATAGCTTTTACGATGGATTATATCTCCCATATGATGCTG 16231

550 sPHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeu 566
:|||||
16232 ACCATCATGCTGCTGTATAGGGGTTTA.....AAGCTGG 16269

567 AlaserLeuGlyLysProLeuHisProLeuValValAlaProAsnGln 583
:|||||
16270 ACGAAGTACTCAGATCGCAGATCAAAGTTGTTTATAACAGCACCTCCACA 16319

583 nLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuH 600
:|||||
16320 ATTGAATTTTGGCTACTAGATACTCAAGGATTGATTATCTACCATTTAT 16369

600 iSHisIleSerMetIleProAlaLysCysLeuGlnGluAlaGluIle 616
:|||||
16370 CCAACATCGTATTTATTTCCAAATAGCTGTTTACGTACTGACAGGAGCGC 16419

617 SerSerProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAs 633
:|||||
16420 TCTGCTTGGATCCTCTAGGTG...TCTTCTTTATTTAAAGAAATTGA 16466

633 pLeuGluGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheG 650
:|||||
16467 TTTAGTATCTTTTCGAACGCTTCTGCCATACATCGCCGTTATTCATATT 16516

650 lYCysAlaLeuValHisThrSerGlyTrpLysValValTyrSerGlyAsp 666
:|||||
16517 GCATGGAGATTACTAATTCACGTGGTTGGAAGATCCCATCTCTGGAGAT 16566

667 ThrMetProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLe 683
:|||||
16567 ACTCGACCTCTCGAGGATTTTCTACATTTGCCAAGACCTCACTATTATT 16616

```

```

683 uIleHisGluAlaThrLeuGluAspGlyLeuGluGluAlaValGluL 700
:|||||
16617 AATTCAGGAAGCTACTTTGGAGACTCTATGCAGAAATGCAATAAAAA 16666

700 ySThrHisSerThrSerGlnAlaIleSerValGlyMetArgMetAsn 716
:|||||
16667 AGCAACACAGTACTTACTCTGAAGCTTTAGAAGTGCCCAAAAAAGCTGGT 16716

717 AlaGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValPr 733
:|||||
16717 ACAAAAAACGTTATTATTGACTATTGACCAACGATATCCCAAAATTACC 16766

733 oLeuPheSerProAsnPheSerGlu...LysValGlyValAlaPheAsph 749
:|||||
16767 AGACATTGATATTTCGACAGAGACTTACATATAGCTTTGGCGTTTCATG 16816

749 iSMetLysValCysPheGlyAspPheProThrMetProLysLeuIlePro 765
:|||||
16817 GAATGACTCTTAAATATCGGATATCTCATTTGTTGTTATTTTGGAAAG 16866

766 ProLeuLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgAr 782
:|||||
16867 CCTCTGGCTTATCTATT.....AATGAGAAATTTAAAGAGGAGTC 16910

782 gGlu 783
:|
16911 GGAT 16914

seq_name: gb_pl1:AC022354

seq_documentation_block:
LOCUS AC022354 46335 bp DNA PLN 01-FEB-2000
DEFINITION Arabidopsis thaliana chromosome I BAC F915 genomic sequence,
complete sequence.
ACCESSION AC022354
VERSION AC022354.1 GI:5850338
KEYWORDS HTC.
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B., Wu,D.,
Bowman,C.L., White,O., Fujii,C.Y., Utterback,T.R., Barnstead,M.E.,
Ronning,C.M., Koo,H., Nierman,W.C. and Fraser,C.M.
Arabidopsis thaliana chromosome I BAC F915 genomic sequence
Unpublished
2 (bases 1 to 46335)
Direct Submission
Submitted (01-FEB-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
Address all correspondence to:atetigr.org

```

BAC clone F915 is from Arabidopsis chromosome I and is near the molecular marker m213.

The orientation of the sequence is from SP6 to T7 end of the BAC clone.

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (available by anonymous ftp from arthur.epm.ornl.gov), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/~chris/GENSCANW.html>), and NetPlantGene (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>), searches of the complete sequence against a peptide database and the Arabidopsis EST database at TIGR (<http://www.tigr.org/tdb/at.html>).

Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are

annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by repeatmasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).

FEATURES

```

source
1..46335
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="m213"
/clone="F915"
misc_feature
1..5778
/feature="overlap with BAC clone F5F19
(AC006216:105116..110893)."
misc_feature
1
/feature="53772 nt beyond this point were not included in the
submitted sequence due to an overlap with another BAC
(F5F19)"
repeat_region
complement(1497..1591)
/rpt_family="(TAAAA)n"
repeat_region
complement(2125..2194)
/rpt_family="(GAAA)n"
mRNA
complement(join(<7284..7610,7681..8652,8869..9016,
9063..9373,9467..9826,9904..9947,10041..10095,
10177..10221,10306..10395,10478..>10639))
/gene="F915.1"
complement(<7284..>10639)
/gene="F915.1"
CDS
complement(join(7284..7610,7681..8652,8869..9016,
9063..9373,9467..9826,9904..9947,10041..10095,
10177..10221,10306..10395,10478..>10639))
/gene="F915.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAF29402.1"
/db_xref="GI:6850339"
translation="MENNEATNGSKSSNSFVFNKRRAAGPDITOKKRNLRKSKDL
NPTNTIAYAIQILGTQDTSSTSVLFFDKQRTFNAGEGLQRCTEHIKLSKIDH
VFLSVSGETAGLPGLLTLTAGIGEGLSVNVWGPSDLNLYLDAMKSEIPRAAMVHT
RSFGSSFPDPLVNVDEVKLSAIIKPCHEEDSGNKGSLSVYVCELPETLIGKGF
DLEKAKVPGKPSYKSLQSGSESDERDITVHPSDVMGSLPFGPIVLVLDQPTF
SHAEFLSKLESYSPDEQITGAKFVNCIIHSPSSVTSPYQSMKKFHLHTQH
ILAGHQRFLPLIIIVSHQKTVKNNAPILKASRIAAARNLYLCPQFPAPGFWPSOL
TNSIIDTPPSKNFLRPAVIRGIDRSCIPAPLTSEVVDELLSIPETIKDKSEERIKO
FNWKGHNTIIEKLISECNTVLPNCLEKIRDDMEIVILGTGSSQPSKYRNVSAIFI
DLFRGSLLDGCEGTGLQKKRYGLDGADEAVRKLRCIWIISHIHADHHTGLARIAL
RSKLLKGVTHPEVIVVGPRLKRFLDAYQRLEDLDMFELDCRSTTATSWASLESGEA
EGSLFTQSPMOSVFKRSDISMDNSVLLCLNKKLVISEIGLNDLISFPVYHCQAY
GVVIAAERNVSGEQILGWKMVYSGDSPETVEASRDATILIHATFEDALTEEA
LAKNHSTKEADVGSAANVYRIVLTHFSORYPKIPVIDESHMHNTCIAFDLMSINNA
DLHLVPLPYEKTIFRDEMVEDADDDVAMDDLKEAL"
10641..10753
/rpt_family="(GAA)n"
10928..11001
/gene="F915.2"
/product="trnA-Ala"
/anticodon="(pos:10962..10964,aa:Ala)
10928..11001
/gene="F915.2"
mRNA
complement(join(11063..11583,12194..>12301))
/gene="F915.3"
complement(11063..>12301)
/gene="F915.3"
/note="similar to delta tonoplast integral protein
GB:AA049281 from [Arabidopsis thaliana]"
complement(join(11317..11583,12194..>12301))
/gene="F915.3"
CDS
complement(join(11317..11583,12194..>12301))
/codon_start=1
/product="putative delta tonoplast integral protein"
/protein_id="AAF29403.1"
/db_xref="GI:6850340"
translation="NAVDEVSIIFIGVASTVSVWVFINKFTYLEAVPTSNAIPIHSVA
VRVGSQTVRMELIITFALVTVTYATAIDSNNGTGLTIAPLAILIVGANILAAGPFS
GGMPMPGRSFGSSLAGVNFSGH"
13176..13409
/note="novel repeat from T25M19 22823 to 23610 788 nt"
13177..13298
/note="novel repeat from T25M19 22823 to 23610 788 nt"
15264..15339
/rpt_family="(GAAA)n"
complement(18843..18918)
/rpt_family="(GAAA)n"
19668..19760
/rpt_family="POLY_A"
complement(20890..20985)
/rpt_family="(GA)n"
complement(20931..20985)
/rpt_family="(GAA)n"
join(<21044..21140,21233..21450,23180..23739,24098..25199)
/gene="F915.4"
<21044..25199
/gene="F915.4"
/note="similar to PEPTIDE TRANSPORTER PTR2-B GB:P46032
from [Arabidopsis thaliana]"
join(21044..21140,21233..21450,23180..23739,24098..25046)
/gene="F915.4"
/codon_start=1
/product="putative peptide transporter"
/protein_id="AAF29404.1"
/db_xref="GI:6850341"
translation="MENPPNETEAKQIQNTNEKKTGGIITMPFIANAEFAKVASYG
LLPNMILIRDYRFQVAKGTNVLFMWSAASNFPLLAGFLSDSYLGRFLTITSIASLS
SFLGVLWLTAMLPOVKPSPCDPTAAGSHGSGSTASQALISAPALISIGSGGRFP
CSLAFGADOLDNKENPKNERVLESFPGWYASSAVAVLIAFTGIIVYIQHGLGKIGFG
VPVLMILIAALLFILLASPLYVTRGVTKSLFTGLAQAIYAAYKKRKLSDPHDHSFDCY
YHMKSEIKAPSKQLRFLNKAACIISNREEEISGDSGFALPWRCLCTDKVEELKALIKV
IPITSTGINMSINTSSSQFLQIATSMRRLSRHSGSFQVPAGSGFMFTIIALAKVI
LYDRAVILPASKIRGRPFRLSLGRMLGLFMSEFLAMASAMVESPRRKAISQGVAN
NSNAVDDISAMMLVPQYVLHGLAEALTATGQTEFFYEPFKMSIIAASLFLGLGMAVA
SLLASVILNAVNELTSRNGKESWSDNIKNGHYNYVWVLAIMSEFINVIYVICSWSY
GPLVDQVRNCRVNGVREEEELDIVGKGEKEDLSIVVKTN"
26156..26179
/rpt_family="(TA)n"
26779..26899
/rpt_family="(CAAAA)n"
join(<26899..26982,27574..>27987)
/gene="F915.5"
<26899..>27987
/gene="F915.5"
/note="predicted by genscan"
join(26899..26982,27574..27987)
/gene="F915.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAF29405.1"
/db_xref="GI:6850342"
translation="WMRYCQFTPYDPSSNMDDLKAPEDETWKYKKKKEAWSLFDEVLY
SEEMKNMGYTLINVMNECFIGRPDQAIQIFVKTALKLKNPDVTCYRNITTLANKQ
GVLEFAEYFFEECMSDRLYPDVFTYTTMIDAYLKAGKTEDALRISNKKWDAFLGQVA
WLACL"
complement(27450..27491)
/rpt_family="(TA)n"
complement(join(<29019..29311,29402..29473,29582..>29780))
/gene="F915.6"
complement(<29019..>29780)
/gene="F915.6"
/note="similar to PGP224 GB:AA02554 from [Petunia x
hybrida]"
complement(join(29019..29311,29402..29473,29582..>29780))
/gene="F915.6"
/note="unknown protein"
/codon_start=1
/protein_id="AAF29406.1"

```


381	381
9004	TTTCCAAATCTCGAAAGCAAGTTCAAGAATCGCTCCAGACTTAACATATC	8955
381	381
8954	TATGCCCGCAGTTCCTTCCCTGCTCGTGAATTTGGCCTTCCTCAACTTACC	8905
382	386
8904	GATAACTCGATTATTGATCCCACTCCCTTCGAACAAGGTTCCTATTTCCTT	8855
387	392
8854	TTCTCATATTAAGTTGAGATAAGTTGATTGTGTCAAATCTAATTAAGTAA	8805
393	404
8804	AGTTCTGATTCCTTATGTCGTCACAGTTCAAACTCTGCTGAACGCAITTC	8755
404	rgCysLys.....	414
8754	TGCTGAAATCTCCTCAAGGTATTGTCATTTCTCCAACACTTTCGCTTTTG	8705
415	424
8704	ACTTTTGAATCTATCTTTAGCTCTGTAAATTTCTCTGATATTTGTGGC	8655
424	styrGlnLeuAsgProAsg.....	440
8654	AGTTCNAITTTGGCTCCGCTGGCTATTCGTGGATCGACAGATCGTGTATC	8605
440	hrcys.....	453
8604	CCTGCTCCACTAACATCTCAGAGTACTGTCAGCTACTTCTCTGAAT	8555
453	uProAsnpheGlnGlnSerValGlnGlytyrArg.....	468
8554	TCCGGAGATMAAAGATAAAGTGAAGAATAACAACAGTCTTGGAAATAAC	8505
468	lnAspGlyProAlaProAlaGlnLys.....	476
8504	ACACACAAGCAATCATCGAAAAGCTATGGCTGCTGAGTCCAATACT	8455
477	484
8454	GTTCTTCCAAACTGTCTTTGAAAAGATTACACGAGATCACATGGAGATTGT	8405
484	ePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSerA	501
8404	TATTTAGACATGTTCTGCTCTACGCTTCAAAATACCGAAACGTTAGT	8355
501	laThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGly	517
8354	CAATTTTATTGACTTGTCTCTAGAGTAGTCTCTCCTAGATTGTGGT	8305
518	GluGlyThrPheGlyGlnLeuCysArgHisTyrGly...AspGlnValas	533
8304	GAAGAAACCTTTGGGCCAGCTTTAAAGAAGATATGCTTTACACGGTGCAG	8255
533	pargValLeuGlyThrLeuAlaIaValPheValSerHisLeuHisAla	550
8254	TGAGCGGGTTAGAAAGTTGAGATGTATATGATCTCTCACATCCACGCTG	8205
550	spHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeu	566
8204	ATCACCATACAGTCTTCCGAGGATTCTAGCTCTTAGAGCAACGCTTTTG	8155
567	AlaSerLeuGlyLysProLeuHisProLeuLeuValValaProAsnGln	583
8154	AAA.....GGAGTCACCTACAGACAGTATGTTGTTGGTGGGCGGAGGCC	8111

583 nLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuH 600
GCTTAAACGCTTTCTTGATGCATATCAAGA.....CTTG 8076
600 ishHisLeSerMetIleProAlaLysCys..... 609
AGATTTAGATATGGAGTTTCWTGATTCTAGAAGCACCACAGCACTTCT 8026
610LeuGlnGluGlyAlaGluIle..... 616
8025 TGGCGCTCTTTAGAGAGTGGGGCGCAACTGAGGTTCTTTGTGTACTCA 7976
617 .SerSerProAlaValGlu..... 622
7975 AGCAAGTCTATGCAGAGTGTCTTCAAAGATCTGATATCTCGATGGATA 7926
623ArgLeuIleSerSerLeuLeuArgThr 631
7925 ATTCTTCGGTTTTCGTGTTTAAAGATCTGAAGAAGTTCTTTAGCGAA 7876
632 CysAspLeuGluGluPheGlnThrCysLeuValArgHisCysLysHisAl 648
7875 ATTGGATTGAATGATTGATTAGTTTCTCTGTGTACATTCTCCTCAAGC 7826
648 aPheGlyCysAlaLeuValHisThrSer..... 657
7825 GTATGTGTGTATCTCAAGCGCATGGAAGAGTAATAGTGTGGAGAAC 7776
658GlyTrpLysValValTyrSerGlyAspThrMetProCysGlu 671
7775 AAATCCTTGGATGGAAATGGTTTATTCCGGCGACTCAAGACCTTGCCT 7726
672 AlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGlu..... 686
7725 GAGACGTCGAAGCATCAGAGATGCAACAATCTTATACATGAGTAAA 7676
686 686
7675 AGTTGTTATTATTAACACAATCTTACAAGTTAGATTTCGACATTCAATACT 7626
687Ala_ThrLeuIleuAspGlyLeuGluGluAlaIaVala 698
7625 GATTGTTTTTGCAGCGCACTGTTGAAGATCGGTTGTATAGAAGAAGCTTT 7576
698 lGluLysThrHisSerThrThrSerGlnAlaIleSerValGlyMetArgM 715
7575 GCGGAAGAATCACACGACACTAAAGAAGCAATCGATCTGGGATCTCGG 7526
715 etAsnAlaGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLys 731
7525 CGAATGTGTATCGGATGTGTTCTCACACATTTTAGCCAGAGATATCCGA 7476
732 ValProLeuPheSerProAsnPheSerGluLysValGlyValAlaIaPheAs 748
7475 ATTCCAGTGATTGATGAGTCGCATATGCATTAATACCTGCATTGCTTTGA 7426
748 phiMetLysValCysPheGlyAspPheProThrMetProLysLeuIleP 765
7425 TCTGATGAGTATAAATCATGGCGGATTTGCATGTCTCTTAAGGTTTTCG 7376
765 roProLeuLysAlaLeuPheAlaGlyAspIleGluGluMetGluGlu 780
7375 CGTATTTCAAACTCTTTTTCAGAGATGAGATGGTTGGAGATGAGGAT 7329

seq_name: gb_pl1:AB023046

```
seq_documentation_block:
```

LOCUS	AB023046	7289 bp	DNA	PLN	19-JUL-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone: MYA6				
ACCESSION	AB023046				
VERSION	AB023046.1	GI:4220645			
KEYWORDS					
SOURCE	Arabidopsis thaliana (strain:Columbia) DNA, clone lib:Mitsui				

```

clone:MYA6.
ORGANISM Arabidopsis thaliana
REFERENCE: Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
AUTHORS: Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
TITLE: Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 3. I.
Sequence features of the regions of 4,504,864 bp covered by sixty
P1 and TAC clones
DNA Res. 7 (2), 131-135 (2000)
20277480
2 (bases 1 to 75289)
Sato,S., Nakamura,Y., Kaneko,T., Kato,T., Asamizu,E. and Tabata,S.
Direct Submission
Submitted (01-FEB-1999) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Department of
Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 252-0812, Japan
(E-mail:ynakam@kazusa.or.jp, Tel:81-438-52-3935,
Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd.graph.cgi?c=MYA6
http://www.kazusa.or.jp/kaos/cgi-bin/agd.graph.cgi?c=MYA6
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli1.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MS11 and the 3' clone is MDC8.
FEATURES
Location/Qualifiers
1..75289
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="MYA6"
/clone_lib="Mitsui pl"
complement(join(483..4904,4991..5106,5192..55636,
6082..6469))
/note="gene_id:MYA6.1
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01260.1"
/db_xref="GI:9279703"
/translation="MERGMNQGSLHWSQSPKTPITMLDRALSSRRRHSADLSASGE
SGTDESKTRPHIVLLASNFLSRIGHOWPCLILALLFLVLLISVAFHSHSEVCIS
RFDPAIRIGFGLDSEDCALGPWCRSKHGEVENTSKDLKGLGEFPIYETRP
IKNNYMGFDHSGNFMFARWLAPDMIESGAPKGSHTWVLRQAMPDTMISLTPRH
PEKYLKRGPAVDGNCITYFAGKEDVFGSDVMKNVLRKHGHTIDLSRVIVFFDDHONEL
RKLQALKAGFRHLIFEDNYDTGCDHYSLRQICDSHIRGGGHSFCRFSDEARMRSK
RKKFWKAVDTFEELCGGETWGWGEMRDRDNTNTPISYNQHFQNSRYVESILDVY
WELPPVAGPSLTHQSRVDPARATPPIVADGKHLRFQIRIGLRDKSVFNGYQTQVYLE
ISKPS"
complement(7397..8479)
/note="gb|AAF25964.1
gene_id:MYA6.2
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01261.1"
/db_xref="GI:9279704"

/translation="MSKFLPBEALTEILVRLSMKDLARPCVCKTWRDLINDPGFTET
YRDMSPAKFVSFYDKNFYMLDVEGKHVPVITNKLDPQLDSMDIDESTCYLHDCGTLCVT
LKNITLWMPNFSQKFIVPNPGIYQDSNIGLFGDVPDHYDVKVTFIDRLDVSIAHV
FEFTSGESLIRSYPDWYDRRGTFDLOYLYWIRSSADRGILCNLSLMSFSFIS
LPLVYNGVTSWMLGVTSQKLCITEYEWMCKEIRISVMEKTSWSKILNLSMSSFSIS
VODRIYDQVEFVSFTRANDLVVFTFTGYNDHFEMEPERTKKKMFPLYKTGNSRSEVR
FCNPLAGURFLCECVETLKIIVNRIFI"
join(10577..10657,10791..10941,11019..11131,11203..11292,
11366..11429,11555..11619,11726..11932)
/note="gene_id:MYA6.3
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01262.1"
/db_xref="GI:9279705"
/translation="MENLNHKWDRCSSSHQTPLSFKGOKOIGOGOREVTFHVSLLPL
AIYDLTKNIEAFONSVLGNKDKPLKFQSTLAEMGIEKSIETVPTKTLTVVMLKLE
NNEVYKQNILOVLSFSEKQSIENVRQALKNRPVFIRLGLSCMSGLDKTRVLYA
PVEEVGHVLIIDAFENVGAGDAKSLRKLHATLMNASYRKDKSKMDTFDAREIHKEF
ENKDWGTYLIIEAHISORYKYDPNGYFHCASLPPPHK"
join(13499..13671,13836..13881,14227..14264,14340..14419,
14556..14623,14732..15011,15103..15215,15295..15360,
15467..15572,15667..15731,15855..15895,15984..16167)
/note="gene_id:MYA6.4
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01263.1"
/db_xref="GI:9279706"
/translation="MIFYQSNKTNQIFFLAWGLEKSFDFVLFLFMSHVVCVGLT
VYLRGVDFVNRNGSEKAMDYKKQKMWGASLIKIRGEGTQTMKLEEMGVKII
LPSSRNKDHISTEGSVDCVTKASKRIATIIDEVVSPSLDYSHEVSLPLAHPLEVL
KLVNFQNSILGHTASDQDQANRATTSVAVDLKANSTNOVNVKTSIPVSYPP
KAKSKSTLLDLGTFKSTFKPSTPHTLVVMLKWNKDRVNAACDLKSIPISSVMDAL
DNKPVFIRLGLDCMRGLDKTRVLYAPVEELGDEGRLLLRACRILDFITDAFVAGL
VLEKDAKQSLKLVHTVMNARKKRNKKNKMETFDAREIHQFGNEDWGETLLOEAF
LSQRFVQNGYRCCGISPIPFGEQRA"
join(18398..18524,18628..18878,19278..19652)
/note="gb|AAC49281.1
gene_id:MYA6.5"
/codon_start=1
/evidence=not_experimental
/product="delta tonoplast intrinsic protein"
/protein_id="BAB01264.1"
/db_xref="GI:9279707"
/translation="MAGVAFGSFDDSFSLASLRAVLAEFISLLEFVAGVGSIAVAK
LTSNALDTPGLVAIVAGCHGAFALFVAVAGANISGHVNPVATFGLVGGQTVITGV
FYTAQLLGSTAAECFLKYVTGGLAVPHTSVAAGLGSIEGVVMEIITIFALVYTVYAT
AADPKKGLGTIAPLAIGLIVGANILAAAGPSGSGMNPARSFGPAAAGDFSGHWYVW
VGPLTGGGLAGLIYGNVFMGSSEHVPLASADF"
complement(join(19955..20047,20150..20212,20303..20365,
20457..20513,20611..20721,20957..21184))
/note="gene_id:MYA6.6
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01265.1"
/db_xref="GI:9279708"
/translation="MGVQLSGSLVASLPPNHSHTKTKLNKPNYSYFRSKHNAART
KTVRAISTAPASQPPADEPDPVAFVHSVLLPDGTPDVHRRRANGGQKLRDIM
LDSNIELYGPSKPLSCNAGVATCATCHEIVHSVLLKLNPRDIDIEKILKRKPKNRL
ACQTNVGNPDSLTGLVIQQLPWEKAHWNIPKNIPNDDDLST"
join(22261..22656,22735..22824,23038..23092,23187..23230,
23347..23742,23833..24098,24182..24326,24415..24465,
24561..25553,25635..25982)
/note="gb|AAF29402.1
gene_id:MYA6.7
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01266.1"
/db_xref="GI:9279709"
/translation="MLTSSMPQNLSLFGFSPLKSSSFALILRPFSLYPPPIFASSSPAP

```


215 rclSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgA 232
| ||||| : ||||
23577 TGGCAGCAAACTGGTGAAACAGCAGTTATATGTATGTGAATTACTT 23595

232 rgGlyValArgAspSerLeuValAlaPheIleCysLysLeuHis 248
:::||| :||||| ::|||
23596ACAGCAGTTATATGTATGTGAATTACTT 23625

249 LeuLysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuPr 265
:||||| :||||| :|||||
23626 GAAATCAAAGGAAGCTTTTGATCCAAAAGAAGGCCATGGCTCTTGATTCAG 23675

265 ovalGlyThrAlaAlaIleAlaProIleAlaValLysaspGlyL 282
:||||| :||| :: :||||| :|||
23676 AGCTGGTCCAAGTAGCATTAATCTTCAGTCTGGAC 23710

282 ysSeriLeThrHisGluGlyArgGluIleLeuAla 293
:||||| :||| :||| :|||
23711 AGTCAGTCAAGTCAGATTTTAAAGATAATTACGGTATGTCCAAAGACA 23760

293 293
.....GlucLeuCystHrProp 300
23811 CACTGTTCCAGCTTAATCATAGTTTCATCGAGTGATGTGATGGCACCTT 23860

300 roASPProGlyAlaAlaPheValValValGluCysProASpGluSerPhe 316
||||| :||| :||||| :||||| :|||||
23861 CAGTTCCCTGGCTCTGTTTGGCTGGGATGTGCACACAGACTGCAT 23910

317 IleGlnProIleCysGluAsnAlaThrPheGlnArgTyrrGlnGlyLysAl 333
::||| :||| :||| :||| :|||
23911 GCCGAGGAAGCTGCTATCTATCCCCTCTATGAAGACTTATTATTCATGCTCT 23960

333 aSPAlapro ValAlaLeuValValHisMeta 344
||| |||| :||||| :||||| :|||||
23961 TGACAATTCAACAGATGGAGCTAGCTTGTAAATGTATTTATTCATCTAA 24010

344 laProAlaSerValLeuValAspSerArGtyrGlnInTrpMetGluArg 360
:||||| :||| :||||| :||||| :|||||
24011 GTCCAGCTCTGTACAAATAGTTCTACATACCGAAGCTGTGATGAAGAGA 24060

361 PheGlyProAspThrGlnHisLeuVal... 369
||| :||| :||||| :||| :|||
24061 TTC...CATTCTGCCACGACATCTTCGCTGGACATGAAGCGTGAGTCTT 24107

370 LeuAsnGluAsn 373
||||| :||| :||||| :||||| :|||||
24108 CTCACATAGTTCTTTCTCATANNAAGACTGTGAATTTATTTGTCCTGA 24157

374CysAlaSerValHisAsnLeuArg 381
||| ||| :||||| :||||| :|||||
24158 TAGTAGTTTTGGTGTGTCGAGGAAGAATATGGAGTTTCCAATCCCTGA 24207

'382SerHisLysIleGlnThrGlnLeuAsnLeuIleHisProaspIle 396
||| :||| :||||| :||| :||| :||||| :||| :|||
24208 GAGCCAGCTTCTAGAATCACGCCAGACTTAATCTTTATGTCACAGTTC 24257

397 PhePro 398
||||| :||| :||||| :||||| :|||||
24258 TTCCCCAGCTCCTGGATTTTGGTCTCATCAACATGATATAACTCGATANA 24307

398 398
..... 398
24308 CCCCACTCTTTGACGAAGTTTACAACGCTTTTCTGATATTTCTCTTTT 24357

399Leu 399
||| :||| :||||| :||||| :|||||
24358 GAGTTCAATCTGGCTGTGTTTCATTTAGTCAAGTTGTGTTCTTTCTCTC 24407

[illegible]

```
25302 ACAGACAATCTTCA...GCTTACCATCTTGAAGAACTG...AAAA 25345
627 rleuLeuArgThrCysAspLeuGluGluGlnThrCysLeuValArgH 644
25346 AGTCCTCGCGGAAATGGGTGGACATTTGATTAGCTTCTGTGTGTC 25395
644 IsCysLysHisAlaPheGlyCysAlaLeuVal..... 654
25396 ATTGCGCTCAAGCCCTTGGGTTTAAAGACGACGAGAAAGAAAAAC 25445
655 .....HisThrSerGlyTrpLysValValTyrSerGlyAspTh 667
25446 ATTGCGGAGATGAGATCCGGGGTGGAAATGGTTTACTCGGAGACAC 25495
667 rMetProCysGluAlaLeuArgMetClyLysAspAlaThrLeuLeuII 684
25496 RAGGCTTGTCTGAGATGGTTGAAGCTTCAAAAGGACACAGTTCTTA 25545
684 leHisGlu..... 686
25546 TACATGAGGTGCTTCTTCTATATAAGCAAAACGTCTGTAACACATTTACATC 25595
687 .....AlaThrLeuGl 690
25596 TCATTTCTTAAATTTACCATCTTTTCAAACTTATTAACAGCGACGTTTGA 25645
690 uAspGlyLeuGluGluAlaValGluLysThrHisSerThrSerG 707
25646 GGATCGGTGTGAGAGAGAGCTGTGGCTTAAACACATAGCACAACTAAG 25695
707 InAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsn 723
25696 AAGCAATCAAACTTGCATCTTCAAGCAGGTGTATACCGAACCGTACTGACA 25745
724 HisPheSerGlnArgTyrAlaLysValProLeuPheSerProAsnPhese 740
25746 CATTTAGCCAGAGATACCCAAAGATCCCGGTGATTGACGAATCACATAT 25795
740 rGluLysValGlyValAlaPheAspHisMetLysValCysPheGlyAspP 757
25796 GCACATACATCATGCTGCTTGATATGATGATGATGATGATGATGATGAT 25845
757 heProThrMetProLysLeuIleProProLeuLysAlaLeuPheAlaGly 773
25846 TACATGCTGCTTCCAAAGATTTTACCTTACTTCAAACTCTATTTCAGAAAC 25895
774 AspIleGluGluMetGluGluArgGluLysArgGlu 786
25896 CAAGTGGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 25934
seq_name: gb_pl3:SCYKR079C
```

```
seq_documentation_block: 3685 bp DNA PLN 11-AUG-1997
LOCUS SCYKR079C
DEFINITION S.cerevisiae chromosome XI reading frame ORF YKR079c.
ACCESSION Z28304.1
VERSION Z28304.1 GI:486556
KEYWORDS baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 3685)
AUTHORS Pohl,T.M. and Pohl,F.M.
JOURNAL unpublished
REFERENCE 2 (bases 1 to 3685)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1994) Data collected by MIPS on behalf of the
European yeast chromosome XI sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitze 18a D-82152
Martinsried, FRG; E-mail: Mewes@hpmc.mips.biochem.mpg.de
```

```
FEATURES
source
Location/Qualifiers
1..3685
/organism="Saccharomyces cerevisiae"
/db_xref="taxon.4932"
/chromosome="XI"
complement(80..2596)
/note="ORF YKR079c"
/codon_start=1
/protein_id="GI:486557"
/db_xref="GI:486557"
/db_xref="SWISS-PROT:P36159"
/translation="MTFTPTSDTKHPLLVQSAHGEKYFFKIGGSGRSRLTE
NKIRISKLKIDIFLTGELNMSDIGLPGMILTIAQDKSNLVLHYGNDILNYIVTWRY
FVFRFGDLNHIKMKVEYKDKIIAIVKSNVKNKNGEDRLGVDSFGKGLVRSIVAK
MPFHAPTDTRYDPSDPHLNVELPDLDAKVEVTNVEISPSVPRGFKVEAEIKLGPV
KGPLFAKLTGQITLIDNGIVTPEOVLENERHFAKVLIDIPDLYLNAFVCKFDY
DCAELGNVYFLGDEVTINDLEAFIDIFEKNNYGVNHHMISHNKISPNITISFEGSAL
TTLKALQVNNLNPKTDRVSDFTDRFDPLSRGTSCKSOEPLNTIIBKDNH
IFSQNTVTTFEPERMEEPKCNINGEADFSWOEIFEESHVKPLEPLADVDVINNO
LHVDNFNNSAEKKKHVEIITLGTGSALPSKRVNVTLVKVPFETDADGNTINRMILD
AGENTLGTTHRMFSQIAVKSIFODLKMVYLSHLHADHHLGIIISVLNEWYKNADETS
YIVVTPQWYHKFVNEWMLVENLEILKRIKVIISCEHFINDSFVRMOTOSVPLAEFEI
LKENSQESNRKLELDRDSSYRDVLIHQWEDLSIEYFOTCRAIHCDAWYSNITFR
MDENNEHTFKVYSIGDTRPNIEKFSLEIGYNSDLLEHATLENLLEDAVKKKHCTI
NEAIGVSNKMNARKLILTHFSQRYPKLPQLDNNIDYWAREFCFAFDSMIVDYKIGISO
ORIFPLINKAFVEEKBEEDVDVESVQDLEVLKHKHKN"
BASE COUNT 1107 a 716 c 629 g 1233 t
ORIGIN

alignment_scores:
Quality: 613.00 Length: 914
Ratio: 1.288 Gaps: 38
Percent Similarity: 52.079 Percent Identity: 25.383

alignment_block:
US-09-434-382-2 x SCYKR079C/rev ..
Align seg 1/1 to reverse of: SCYKR079C from: 1 to: 3685
29 ArgGluArgProArgLysAspProLeuArgHisLeuArgThrArgGluLys 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2671 AGAAGAGGACCTTAAAGAACACAGACTAATAGACGACACCCGCAAGA 2622
45 sArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrLeuGlnV 62
||||| ||||| ||||| ||||| ||||| ||||| |||||
2621 AAAGAGTCTCGAAGGTTTCA.....ACGATGTTACATTTATACCA 2578
62 alValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrValPheSer 78
||||| ||||| ||||| ||||| ||||| ||||| |||||
2577 TCACCCATCTACATCGGATACAAAGCACCCATTCGTCTAGTCAGTCT 2528
79 GluPhe....AsnArgTyrLeuPhe....AsnCysGlyGluGlyValGlnAr 93
||||| ||||| ||||| ||||| ||||| ||||| |||||
2527 GCACATGGGGAAGATATTTCTCGTAAATTTGTTGAAGGATCCCAAG 2478
93 gLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePheL 110
||||| ||||| ||||| ||||| ||||| ||||| |||||
2477 GAGTCTGACTGAAATAAGATCAGATATCCAAATTTGAAGGATATTTCC 2428
110 euThr...ArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
||||| ||||| ||||| ||||| ||||| ||||| |||||
2427 TTACTGGTGAATTAACCTGGTTCAGATATAGTGGATACCTGGAATGAT 2378
126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProPr 142
||||| ||||| ||||| ||||| ||||| ||||| |||||
2377 TTGACTATTGCTCAAGGGAAGAAATCTTGTTTTTCATTACGGCAA 2328
142 ogLnlLeuGluLysTyrLeuAlaIleLysIlePheSerGlyProLeuL 159
||||| ||||| ||||| ||||| ||||| ||||| |||||
2327 TCACATTTTGAATTACATAGTTTCCACCTGGAGATACCTCGTCTTTAGAT 2278
159 ysGlyIleGluLeuAlaValArgProHisSerAlaProGlu..... 172
```

1494 TCACGAGAGGTACATCTATGTGTAAATCCCGAAGAAGCCTTTGAATAACA 1444

433TrrpGlnArgAspAlaIleLeuTh 440
|||||:||||
1444 ATAATAGAGAAGGATAACATTCATATTTCACAAAACAGACAGCATAC 1395
|||||:||||
440 rCysAsnPro..... 443
|||||:||||
1394 TTTTCAACCATTTCGGATGAACGAAGAACCGTGAAATGCAACATCAACG 1345
|||||:||||
444GluGluPheIleVal 448
|||||:||||
1344 GTGAAGTGGCGGATTCCTCGTGGCAAGAAATTTTCGAAGAACAT...GTA 1298
|||||:||||
449 GluAlaLeuGlnLeuPro.....AsnPheGl 457
|||||:||||
1297 AACACATTAGAAATTCCTTAGCTGATGTCGATACACTTATCAATAATCA 1248
|||||:||||
457 nGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlaProA 474
|||||:||||
1247 ACTACAGCTGGATAACTTTAACAAATTCAGCAGAA..... 1214
|||||:||||
474 laGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySer 490
|||||:||||
1213AAGAAGAACACGCTTGAATATTCACCTTAGGAACCGGTAGT 1172
|||||:||||
491 AlaIleProMetLysIleArgAsnValSerAlaThrLeuValAsnIle.. 506
|||||:||||
1171 GCATTGCCTTCTAATATAGAACCTTGTCTCCACACTTGTGTAAGTTC 1122
|||||:||||
507SerProAspThrSerLeuLeuLeuAspC 516
|||||:||||
1121 TTTTACTGACGCCGATGGAATACCAATAAGTAAGAACATATGCTAGATG 1072
|||||:||||
516 ysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAsp...Gln 531
|||||:||||
1071 CTGCTGAAAACTATTTAGGTACCATCACACAGGATGTTTTCTCAGCTAGCA 1022
|||||:||||
532 ValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHi 548
|||||:||||
1021 GTCAGTCAATATTTTCAGGATTTGAAAATGATATATCTGAGTCACTTGCA 972
|||||:||||
548 sAlaAspHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgA 565
|||||:||||
971 TGCAGACACCATTTGGGAATAATCAGCGTGCTA.....AATGAATGGT 928
|||||:||||
565 laLeuAlaSerLeuGlyLysProLeuHisProLeuLeuValValAlaPro 581
|||||:||||
927 ACAATATAACAGGATGATGAACAGGATATATATATATGTTGTTACTCCA 878
|||||:||||
582 AsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsn..... 593
|||||:||||
877TGG.....CAATATCAACAAATTTGTTAATGAATG 849
|||||:||||
594GlnCysGlnGluValLeuHisHisIleSerMetIleProA 607
|||||:||||
848 GTTAGTTCTAGAAAAATAAGAGATTTTAAAGAGATCAATACATAGTT 799
|||||:||||
607 la..... 607
|||||:||||
798 GTGAGCATTTTCATCAATGATTCGTTTGTGAAGATGCAGACACATCTGTT 749
|||||:||||
608LysCysLeuGluCluclyAla..... 614
|||||:||||
748 CCTTTGGCAGAGTTCAATGAAATATTGAAGAAATAGCAATCAAGAATC 699
|||||:||||
615GluIleSerProAlaValGluArg.....L 624
|||||:||||
698 AACAGAAAACTGGCACTGGATAGAGATTTCTCATATAGGGATCTTCACT 649
|||||:||||
624 euIleSerSerLeuArgThrCysAspLeuGluGluPheGlnThrCys 640
|||||:||||
648 TGATCAGACAAATGTATGAGGATTTATCGATAGAAATATTTTCAAACCTGC 599
|||||:||||

641 LeuValArgHisCysLysHisAlapheGlyCysAlaLeuVal..... 654
:: ||||| |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
598 AGAGCTATACATTGTACTGGCGATAATTCGAACCAATACCTCCGAAT 549

655HISThrSerGlyTrpLysValValGlyrSerGlyA 666
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
548 GGACGAAACAATGAGCATATAACA...TTCGAAGTTTCATATTCAGCGC 502

666 spThrMetProCys.....GluAlaLeuValArgMetGlyLysAspAla 680
||||| ||| ::::|:::|:::|:::|:::|:::|:::|:::|:::|
501 ATACAAGACCTAACATCCAGAAAATTTCCCTCGAAATAAGGCTAATTC A 452

681 ThrLeuLeuIleHisGluAlaThrLeuGluAspGlyLeuGluGluAla 697
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
451 GATCTATTAATTCACGAAGCTACACTAGAAAATCAGCTACTGAGGAGTG C 402

697 aValGLULystrHISerThrThrSerGlnAlaIleSerValGlyMetA 714
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
401 CGTGAAGAAAAACACGCACCTATTAAATGAAGCAATCGGTGTTCGAAC A 352

714 rgMetAsnAlaGluPheIleMetLeuAsnHisPheSerGlnArgTrpAla 730
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
351 AAATCAATGTAGGAAGTTGATCTTAACACACTTTTCCAGAGATATCC C 302

731 LysValProLeuPheSerProAsn.....PheSerGluLysValGl 744
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
301 AAATTGCCCCAATTAGACAATAATATTGATGTGATGGCAGAGAATTTTG 252

744 yValAlaPheAspHisMetLysValCysPheGlyAsppheProThrMetP 761
||||||| ||| ||| ::::|:::|:::|:::|:::|:::|:::|:::|
251 TTTTTCCTTCGACAGTAGTCGTTGATTATGAGAAAATCGTGAAACAGC 202

761 roLysLeuIleProPheLeuLysAlaLeuPheAlaGlyAspileGluGlu 777
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
201 AGCGTATTTTCCACTGCTGNAATAGGCCATTT.....CTTTGAAAGA 161

778 MetGluGluArgGrgGluLysArgGluLeuArgGlnValArg 791
||||:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
160 AAGGAAGAAGAANGATGTTGATGACGCTTCAAAGCGTPACAA 119

seq_name: qb_pl1:AF188714

seq_documentation_block:	
LOCUS	AF188714 19158 bp DNA 31-JAN-2000
DEFINITION	Emericella nidulans chromosome VIII cosmid SW06H01, genomic sequence.
ACCESSION	AF188714
VERSION	AF188714.1 GI:6707311
KEYWORDS	.
SOURCE	Aspergillus nidulans.
ORGANISM	Aspergillus nidulans. Eukaryota; Fungi; Ascomycota; Eurotiales; Trichocomaceae; Emericella.
REFERENCE	1 (bases 1 to 19158)
AUTHORS	Akeno,T. and Akita,O.
TITLE	Direct Submission
JOURNAL	Submitted (22-SEP-1999) National Research Institute of Brewing, Kagamiyama 3-7-1, Higashi-Hiroshima, Hiroshima 739-0046, Japan
REFERENCE	2 (bases 1 to 19158)
AUTHORS	Gomi,K.
TITLE	Direct Submission
JOURNAL	Submitted (22-SEP-1999) Department of Agricultural Science, Tohoku University, 1-1 Tsutsumigori-Amamiyamachi, Aoba-ku, Sendai 981-8555, Japan
REFERENCE	3 (bases 1 to 19158)
AUTHORS	Kunihiro,S., Nakagawa,M. and Machida,M.
TITLE	Direct Submission
JOURNAL	Submitted (22-SEP-1999) Molecular Biology Department, National Institute of Bioscience and Human-Technology, Higashi 1-1, Tsukuba Ibaraki 305-8566, Japan
FEATURES	Location/Qualifiers
source	1..19158

```

/organism="Aspergillus nidulans"
/db_xref="taxon:5072"
/chromosome="VIII"
/clone="cosmid SW06H01"
/note="overlaps cosmid SW06E08 deposited in GenBank"
Accession Number AC000133"
BASE COUNT      4815 a  4709 c  4726 g  4908 t
ORIGIN

alignment_scores:
    Quality:      610.00      Length:    1183
    Ratio:        1.180      Gaps:      42
    Percent Similarity:  43.702      Percent Identity:  22.316

alignment_block:
US-09-434-382-2 x AF188714/rev

```

```

alignment_scores:
  Quality: 610.00      Length: 1183
  Ratio: 1.180         Gaps: 42
  Percent Similarity: 43.702  Percent Identity: 22.316

alignment_block:
  US-09-434-382-2 x AFL188714/rev  ..

  Align seg 1/1 to reverse of: AFL188714 from: 1 to: 19158

      7  LeuLeuArgSerAlaAlaGlyArgThrMetSerGln.....G1 19
      18045  TTACTTCACGTGCGCCATCCGTCACATGTACAAATAAAGAGGCGCT 17996
      19  YargThrIleSerGlnAlaProAla.....ArgArgGluArgp 32
      17995  TCGGCATACTCTACAGCTCCGCAGCTCGATAGTCGGCTCCCTCAAGC 17946
      32  to...ArgGlyAspProLeuArgHisLeuArgThrArgGluLysArgGly 47
      17945  CCTCGGAGAAAGACAATTCGAAACATTATACGAAGAAGCAAAAGCGA... 17899
      48  ProSerGlyCysSerGlyGlyProAsnThr..... 57
      17898  ....TCTTCGCTCTACAATACATCACCTATTCTGGAACACACC 17861
      58  .........ValTyrLeuGlnValV 63
      17860  CCAAAATCCAAAATTCGGAATTACGCAAAATGAAGTTCTACTATCAGATCT 17811
      63  aAlaAlaGlySerArgAspSer...GlyAlaAlaLeuTyrValPheSer 78
      17810  TCAGACGCCAGCGGTGATACCCAGGCTCGGGCGCTTCTCCTGCATTTC 17761
      79  GluPheAsnArgTyrLeuPhe...AsnCysGlyGluGlyValGlnArgLe 94
      17760  CAGGGAACGGTTATCTTTTCGGCCAAATTCGAGGGGTTCCACACAGC 17711
      94  uMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePheLeuT 111
      17710  ATGTACGACACAGGAGGACAAAGTTGACGGACGTCTCATGTTTTCCTGT 17661
      111  hr...ArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIleLeu 126
      17660  CGGTCGGATGGGATGGGACACAACTGGAGGCTTCATTGGCATGATTCTT 17611
      127  ThrLeu..... 128
      17610  ACGAAGGCCGATGCTCGGGCCAGTTCAAAGAGGCTCTGGAAGCCCTCGA 17561
      129  ....LysGluThrGlyLeuProLys..... 135
      17560  GCGGAAAAGGAGGCGAATCCACAGACGAGGCTGGACCAAGAAAGATA 17511
      136  ....Cys 136
      17510  CGAAACAAACCGCCCAACACGAGAGCTCGAAAGATCGCGACGACGACTTG 17461
      137  ValLeuSerGlyProProGlnLeuLysTyrLeu..... 148
      17460  ACTGTCACGGTTCAAGAATCTTTCGTCATACCCCTGGCTACTGCGCGGAG 17411
      149  ....GluAlaIleLysIlePheSerGlyProLeuLysGlyI 161

```



```

17410 GYTTGCTTCAGACAGGTTTCCCGTGCATACGAAGGAGTACGACGCCG 17361
      ::::::::::: :::: ::::
161 leGluLeuAlaValArg.....ProHisSerAla 170
      :::::||||| ||| :::::
17360 AAAGTGTTCAAAGAGGCTGCGCACATGATCAAAGCGACCCCTTTGAAGTG 17311
      :::::||||| ||| :::::
171 ProGluTyrGluAspGluThrMetThrValTyrGlnIleProIleHis 187
      ||| :::::||||| ::::: ||| ::::: ||| :::::
17310 CCGACTTTTCGGATGAAACATCAAGGTGACCATCTGACCATCTATCAGCCC 17261
      ::::: ::::: ||| ::::: ||| ::::: ||| :::::
187 rGluGlnArgArgGlyLysHisGlnProTrpGlnSerProGluArg.... 202
      ::::: ::::: ||| ::::: ||| ::::: ||| :::::
17260 GTCTTCTACTACACCAAGTTCC.....CAAGCCCTAAAGAGGA 17220
      ::::: ::::: ||| ::::: ||| ::::: ||| :::::
202 ..... 202
17219 GTCTTCACGAATTTCCGAAGATGTGCGGGTTTAGTGGAGTGGACCAG 17170
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
202 ..... 202
17169 CAATCCAAGGATCAATTTGGTGGCGAGCTGTGTTGTCTCATATGTTGA 17120
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
203 .....ProLeuSerArgLeuSerProGluArgSerSerAsp..... 214
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
17119 CTCAACTTGGACTCTGGATGCATTTGGAGGAAACCCGTCTTGGCCGAGCTCA 17070
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
215 .....SerGluSerAsnGluAsnGlu 221
      ||| ::::: ||| ::::: ||| ::::: ||| :::::
17069 AATGCCAGCGCAAACTTTGTACGAATCCGGAGACCAAGGACCTGGAC 17020
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
222 ProHis.....LeuProHisGlyValSerGlnArgArgGlyValAr 235
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
17019 AAGTACACAGCGCCGCTTCTCGGAGGAGATGAGGAG.....GTTC 16979
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
235 gAspSerSerLeuValVal..... 241
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16978 TGAGATGACCGTTCGTTCGTGAACCCCTGGCCCGCGCTCGATCAGCA 16929
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
242 .....AlaPheIleCysLys 246
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16928 AGATCCCAACAACACACCATCGTTGAGAGTCTATCTTACATCATCAAG 16879
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
247 LeuHisLeuLysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGl 263
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16878 AACCCAGACCTCGCGGAAAGTTGTGCTAAGAAGCAATCGCTCTGAA 16829
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
263 yLeuProValGlyThrAlaAlaIleAlaProIleIleAlaValLysA 280
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16828 TGTCAACCCCTGGC.....CCTGACTTTGGTGTCTGCACAA 16794
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
280 spGlyLysSerIle...ThrHisGluGlyArgGluIleLeuAlaGluGlu 295
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16793 GGGGTGAACCTGTCAAGGTACTACTGTGAACGATGTGTGACACCTGAATG 16744
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
296 LeuCysThrProAspProGlyAlaAlaPheValValValGluCysPr 312
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16743 GTCTGGAGCCTCAAAAGCCTGGCAAGGACTGGCAGTAATGACCTCCC 16694
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
312 oAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheGlnArgT 329
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16693 GACTTCGGATTATGTGAGAGCCTGTGTAAACCCGCCCGCCGAATGGAAG.... 16648
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
329 yrGlnGlyLysAlaAspAlaProValAlaLeuValHisMetAlaPro 345
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16648 ..... 16648
346 AlaSerValLeuValAspSerArgTyrGlnTrpMetGluArgPheGl 362
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16647 ...TCCCCCTCAGTTGCAAGCAATTTGGCAGCTTCTTTTGGATCCTGGG 16601
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
362 yProAspThr....GlnHisIleValLeuAsnGluAsnCysAlaSerValH 378
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::

```

```

16600 TCCTGGTGTGCGGGAACACACCTCGGCTTCGGGAATTCGTGCTTCATATGC 16551
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
378 isAsnLeuArgSerHisLysIleGlnThr..... 387
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16550 CGAACTGCAAA...CACATCGTCTCGAGCAGCGGATTATTGCCCGAATAT 16504
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
388 .....GlnLeuAsnLeuIleHi 393
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16503 CTGACCATGCAGAGTGTGCGGGTCCGCCATTCGCATGCCCGGATAAG 16454
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
393 sProAsp.....IlePheProLeu..... 399
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16453 GCTGATAATTACCTAGTCTCCTATACAGCAACAACAGTGTTCACACAAA 16404
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
400 .....LeuThrSerPheArg 404
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16403 CGCAATCCAGTCTGGAAGATCGCATGSCCGTGGATAGCCAAACCATTTCA 16354
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
405 CysLysLysGluGlyProThrLeuSerValProMetValGlnGlyGluCy 421
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16353 TTCGAACGAGCCGAGCCCGGAGCTCATGCTTAGTATG...GAGCCCGAGTT 16307
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
421 sLeuLeuLysTyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaI 438
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16306 CAAGATAAACTCAGATCAACTACCGCAGCA...CTGAACGCGAGGTAGCA 16260
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
438 leIleThrCysAsnProGlu.....GluPheIleVal 448
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16259 TCCTGACAAAGATCCCGACAGTCTGCTAAGCGGGCTCGACTAGTTACC 16210
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
449 GluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgAr 465
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16209 AAGAGGTGCAACATCCGCTACGCCAGGAGAAGATTTCAGCAATTTGTGAG 16160
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
465 gSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProG 482
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16159 CGAATTCCTGGTCCGATCGC.....G 16137
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
482 luIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsn 498
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16136 AGATCATCACGCTCGGAACGGATCTTCGCACCTTCCAAATATCGGAAT 16087
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
499 ValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAs 515
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16086 GTTTCAGCAGCCCTGTTTCATGTTCCAGGTGGGGATATTATCTATTGTA 16037
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
515 pCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGly...AspG 531
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
16036 TTGTGGCGGGCAGCGTGTGTCAGCTTAAGCGCATGTTTCAGCCCTCAGC 15987
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
531 InValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeu 547
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
15986 AGCTGCGTGAAGTTCTCAAAACCTCGCGCTAATCTCGATTAGCCATCTC 15937
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
548 HisAlaAspHisHisThrGly..... 554
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
15936 CATCGGAGCCACCCACCTTGTTACAGTCTCTGTTATAAGGCTTGGTACCA 15887
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
555 .....LeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerL 569
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
15886 GGAAGAACTATCCCGGAGGTGTCACAGCTGAGCGAGCCCTCAAAAGACA 15837
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
569 euGlyLysProLeuHisPro.....LeuLeuValValAlaProAsnGln 583
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
15836 TAGGGAAGATCCTGGAGGAGAGCGTTTGTGTCGTATCTGATCTCATG 15787
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
584 LeuLysAlaTrpLeuGlnGlnTyr..... 591
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
15786 ATGATTGAGTGTGGAGGAATATGCTGGTGTGGAAGATTTTGGTTTCGC 15737
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::
591 ..... 591
15736 CAAGCTCACCCCTCTATCTGCCTATCTATACAGACTGGCAGAGGGGA 15687
      ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| :::::

```

592HisAnGlnCysGln..... 596
15686 TGAATGTCATTCAAAATACCCCACTGCCGTGCCACGCTCTTATCCT 15637
597 ...GluValLeuHisHisHisSerMetIleProAlaLysCysLeuGlnG1 612
15636 GGGCGGGAGTTGACACATATAA.....CCGA 15608
612 uGlyAlaGluLe.....SerSerProAlaValGluArgL 624
15607 ACAAAGTGAACCTGAATTAACATGCAAGGATCACAGCA...GAGGAGC 15561
624 euileSerSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCys 640
15560 TAAGCGCAAGCTCCCAAGCCACGGCTCTGTCGATATTTCAACGGCA 15511
641 LeuValArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSe 657
15510 TAGCTCTCCATTCGCGTGGTGCAATGGAGTCTCACTCATCTTCCCTGA 15461
657 rGlyTrpLysValValTyrSerGlyAspThrMetProCysGluAlaLeuV 674
15460 TGGCTTCAAGTCTCTCTCAGGTGACTGCCGGCCATCGCTAGCTTG 15411
674 alArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGlu 690
15410 TGACTATTGGACAGACACTACCGTCTTATCCAGCAAGCCACATTTAGC 15361
691 AspGlyLeuGluGluAlaValGluLysThrHisSerThrThrSerG1 707
15360 GATGACATGGTGGCTCGCTCTCCCAAAAACATTCACCGCCCAAG 15311
707 nAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMetLeuAsnH 724
15310 AGCTATTGAAGTAGCCGCAAAATCGCGCTCGCACTATCTCTCACC 15261
724 isPheSerGlnArgTyrAlaLysValProLeuPhe..... 735
15260 ACTTCAGCAGCGCTACCAAAAATCGCCCACTTCAATCAGCCCAAGAA 15211
735 735
15210 CTCGTCCAGGCAAGTCACTAGTCACAGCCCGTCACTTCAGGGCACACAAAG 15161
735 735
15160 TATCAGCGCAAGCAAGAAACAGCAGCAGCGGAGCAGCATCCACAGATA 15111
735 735
15110 TTCCCTTCAAGAACGACATTTGATGGCGCAGAGAACATTTGCCCTTGATGAA 15061
736SerProAsnPheSerGluLysValGlyVal..... 745
15060 CTGCGGCTTGAACCCGAGCCAGCCAGGAGTGGTTAGAACATATCTGAGCC 15011
745 745
15010 TTCAGAGCTTCCAGAACCGGATGCCACAGCAGCTCCCATCGTCCCATAA 14961
746AlaPheAspHisMetLysValCysPheGlyAspPheThrMet 760
14960 TTGCAGCATTTGACACATGCGCTCCGCGTCCGCGATATGTACACACTC 14911
761 ProLysLeuIleProLysAlaLeuPhe..... 771
14910 GAACAATATGACCTGCGTTGAGCGCTCTTCGATATCATCATGAGCGGC 14861
772AlaGlyAspIleG 776
14860 CTCCAAACTCGAACAAGCAAGCGGAGAAAGCGAGCGGTGAGGTGC 14811

776 luGluMetGluGluArgArgGlu.....LysArgGluLeu 787
14810 AGGAAACAGAAAGAGGAGTTTAGCGGGAGAACAGTTTAAACA 14761
788 ArgGlnValArgAlaLeuLeuSerArgGluLeuAlaGlyGlyLeuCl 804
14760 AGCAGGAGCAGATGAGCAGGAGCAAGGAATGGCAAGCGGAGAA 14711
804 uAspGlyGluProGlnGlnLysArgAlaHisThrGluGluProGlnAla 820
14710 GGAAGCCCGGTGCTGATAGTCTTCTCTCAGTCTCGGCTCAATCT 14662

seq_name: gb_in2:CELE04A4

seq_documentation_block:

LOCUS CELE04A4 22846 bp DNA INV 19-DEC-1997
DEFINITION Caenorhabditis elegans cosmid E04A4.
ACCESSION AF038611
VERSION AF038611.1 GI:2702410
KEYWORDS

SOURCE
ORGANISM
Caenorhabditis elegans strain=Bristol N2.

REFERENCE
AUTHORS
Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M.,
Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N.,
Latrelle,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra.A.,
Saunders,D., Showkneen,R., Smaildon,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Sproat,J. and Wohldman,P.

2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Nature 368 (6466), 32-38 (1994)
94150718
2 (bases 1 to 22846)
Sammons,L., Wohldmann,P. and Biewald,T.,
The sequence of C. elegans cosmid E04A4
Unpublished (1997)
3 (bases 1 to 22846)
Waterston,R.
Direct Submission
Submitted (15-DEC-1997) Department of Genetics, Washington
University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rw@nematode.wustl.edu and jesus@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is K08F11, 200 bp overlap; 3' cosmid is T16C6. Actual start of this cosmid is at base position 197 of CELE04A4.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES
source

Location/Qualifiers
1. .22846
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="IV"
/clone="E04A4"
3901. .6015
/gene="E04A4.3"
join(3901. .3962, 4015. .4289, 4457. .4560, 4620. .4746, 4793. .5072, 5424. .5610, 5658. .5768, 5815. .5921, 5967. .6015)
/gene="E04A4.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB92038.1"
/db_xref="GI:2702414"

gene

CDS

/translation="MDTAKWFTCFEWEQVIREKKYLEATQVITRELEMLKEDKKLTL
RVPACNDKLPDYTCRSKQKMGDHHVVRPEPVYKQQLFQRTTHVKCTQFETGWIT
RHVELLDKLAITYTAVLKTQDVPHRTIRGALKVARDNDIPITSKOMQNFGRHPTAVP
KGSPPRATTVEHLEKLELEDEVMFSIDNSNLAETIYKHFFDAKMLTAGPTEET
ELDWRDEVERICMDSDTSRSALKILLERTDTRKARIFLDSTFELTDGYTVVVM
GELNSFTKTRVPIPLAYMIHTDLATTHAFANWLTDEKNAIGMGRRRCPLLV
DGEAALVYAEKLTDLVRCDVHIMSLMQHKYKKVLDSDRKLFLFEGCRNTWARGLL
GSFEADFDSEKLVETKSPNSDVWRNRMFLILNASVL"
8026. .9173
/gene="E04A4.2"
join(8026. .8046, 8464. .9058, 9115. .9173)
/gene="E04A4.2"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB92037.1"
/db_xref="GI:2702413"

gene

CDS

/translation="MLLKIEIRGLSLDDPLTFEMFDGNGFKFTTRIFONLKGSYVO
KGLTFLQSDSVLIANPDPDIOMNAVLCNLHETSLKWFRAQPEDLLISNRYL
SISEDMSKKVNLKHLKHWGNSKRMFYDVCVRVFNDDVLAFLFKNDICTAPVE
RELIVQVPEGEQIVKGYDVRMDGTATVIIVDHIOSLKMFVRVNEYIITIIII
YVLEEA"
10829. .12409
/gene="E04A4.1"
join(10829. .10920, 11090. .11970, 12153. .12409)
/gene="E04A4.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB92036.1"
/db_xref="GI:2702412"

gene

CDS

/translation="MVDVSVFPOLLKLPKSSNVVHRMLVADLIGFSLTSKNAKLO
TKRLGLKAEDVHLEIRSGIIHISLOHTFTFYKTYNEYRSEHLEBECINMYLCTR
LGKWTFLYVRQKDLVDLDFHVNINISELEITTDHFTTYDCKLEIEISFSWYHI
DLNKLTQFVRSKLRANQSLILMSRNPTAETSQKQVVLTRNMDHIVLYHPKIYD
LNDLVNTSLYMDMKIEDVMELKHWIGDLKLELACFGKYLKAEVAVELFL
KNVDYQVAPARELTIIPTGSOQICVKGVDVRRVDGEENITVNSORKCLTKIPGSL
FESVHFYCVCATNQRNLLYPLEFIKIMKIYTSNAEKSFDAHNTETPKNSIKHERRD
PRPIENVPFNLLIGIG"
complement(12921. .17200)
/gene="E04A4.4"
complement(join(12921. .13023, 13237. .13460, 13999. .15037, 15263. .15402, 15449. .16319, 16572. .16703, 17175. .17200))
/gene="E04A4.4"

gene

CDS

/note="contains similarity to sulphatases; coded for by C.
elegans cDNA yk19h5.3; coded for by C. elegans cDNA
yk39c9.3; coded for by C. elegans cDNA yk15d11.3; coded
for by C. elegans cDNA yk42h12.3; coded for by C. elegans
cDNA yk22g6.3; coded for by C. elegans cDNA yk19h5.5;
coded for by C. elegans cDNA yk22g6.5; coded for by C.
elegans cDNA yk15d11.5; coded for by C. elegans cDNA
yk42h12.5; coded for by C. elegans cDNA yk39c9.5"

alignment_scores:

Quality: 591.50
Ratio: 1.259

Length: 1038
Gaps: 35

/codon_start=1
/protein_id="AAB92042.1"
/db_xref="GI:2702418"
/translation="MKMLFFGKIVSRHLISSTCLCFKONNELELSEIKERIAARRIL
QKHSSHLKAREVNASISLRQSMMAVOKKOKAHEPPANSIVNIPSOVSIEVLNGT
GLLRACFLTRPLKTYMPCNACRFQWLRIRSSVVDLFTISANWNTAGISIL
LSKSNLPLRLHGMNKKHLECIKRFQDSDYGSCKYPSQVEERTYMENTEDAGLK
VTYIPSLNIGSNMKNVKNVNDIAFLIEMKEARADITMKLMELKVPKGPLI
GKLKSGEAVTLPDGRTIQPDQVFSDDKGVKPLLLVTECTEDHVKALIDSSSLQPF
LNGEKQDVMVHISDDAVNTPTYRHLMEKLNPSITILLINGNPVTPAVESVYKHT
RLLSRIASPLPALHPIDWSGIITONEELSORQDQIRVAPMQRVMMRGASNFEEPI
VNNLLAEPESLSDKAKELIKYQKLEKENKMDCEPKLTFCTSSAYPSKYRNVGYL
VEASNSAILDVGEGTIGQMRVFEFGDQKOLLVNLCLVLIHAHQDHMNGLYTIIA
RRKEAFESLGAIPRLVLCNNRNLKPMKTYSCIFENIEHLEIIVDISRYPLTPPSP
GGPGKRRPLSPHLPSPDRVLDQMSSEDFKAKWLDELKAVQVHITRMANGFVRVA
GKRIVSGDTKPCDLLVEEGDADVLVHESFDEGDHEVMDPKPKKLAKISLADAM
RKRHVSQAVDVGKRMNAKHILLTHFSARYPKVPVLPVLDKENGKIGVMDMLRVRED
HLPLYSKLLPIPREVFVAELFELTIKKEQRLVKDKELSEKROQLKA"
complement(17983. .18738)
/gene="E04A4.5"
complement(join(17983. .18093, 18151. .18261, 18318. .18503, 18553. .18664, 18713. .18738))
/gene="E04A4.5"
/note="contains similarity to S. cerevisiae mitochondrial
import protein (MIM17) (GB:X87371); coded for by C.
elegans cDNA yk83e7.5; coded for by C. elegans cDNA
yk211c5.5; coded for by C. elegans cDNA yk32a7.5; coded
for by C. elegans cDNA yk465g9.5; coded for by C. elegans
cDNA yk81g5.5; coded for by C. elegans cDNA yk255e2.5;
coded for by C. elegans cDNA yk11e10.5; coded for by C.
elegans cDNA yk400c6.5; coded for by C. elegans cDNA
yk452a5.5; coded for by C. elegans cDNA yk452a5.3; coded
for by C. elegans cDNA yk211c5.3; coded for by C. elegans
cDNA yk502c12.3; coded for by C. elegans cDNA yk465g9.3;
coded for by C. elegans cDNA yk32a7.3; coded for by C.
elegans cDNA yk11e10.3; coded for by C. elegans cDNA
yk83e7.3; coded for by C. elegans cDNA yk255e2.3"
/codon_start=1
/protein_id="AAB92039.1"
/db_xref="GI:2702415"
/translation="MEETREPCPYRIGDDIGSAFAMGLVGSIFQAFGGYKNAAGK
KLVGMRREVRMRSTLTGVOFAAGMFSTDDCLVAIRKKEDPINSIVSGGTLGALLA
IRSGPKVWAGSAILGSLVILAMTEGQVLVTRMMGAMMDPTOPPEALDDPRSLGQSK
AEPGLDQFRPGIPTGLPNLS"
complement(18757. .18758)
/note="5' UTR splice site; see EST yk83e7.5"
complement(19059. .21097)
/gene="E04A4.6"
complement(join(19059. .19084, 19138. .19318, 19365. .19496, 19547. .19794, 19849. .20377, 20760. .20863, 20917. .21097))
/gene="E04A4.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB92040.1"
/db_xref="GI:2702416"
/translation="MGTAARNRNLVHLLGLFLLTYCSADTPEVCYPSVWRHRI
DSIFPTSLFCVSKNDLVDGDIVGEOWNEFESKRWQDSDSDSENLLEHFG
SSVRLKFKFPAAISASDECTGQNTVITKRNMTVKCKPLATDILLHSCSENEFLHQA
LFGNGDFVERTENETIAPRLSTAQFIAPVWNSMVCYGLKANSIIRNMGIRKINSQA
IDAQVDSLPGNVNDNNFQTFVSVMIPVIVASEIETNTVGVKAGEQYRVKGSIPV
PFAVTLGNCYSGSIAPSVFFLRSMSSVCTISTINCEADAKARAFYBQVYPSVLVS
SPSEDAHPASVGRVNVNTPWELSPSTISCRLLPVSSLLQIYVSKQSGTKNREVITAGNS
QLLLDDVPLVSGQEIRMPISISFTEVTPPKNIFSAIPVDIRLPHDFPFPMTSINY
SASNYSQLFILVNEY"
complement(21320. .21706)
/gene="E04A4.7"
complement(join(21320. .21475, 21527. .21706))
/gene="E04A4.7"
/note="C. elegans cytochrome c (SW:P19974); coded for by

Percent Similarity: 45.279 Percent Identity: 21.484

alignment_block:

US-09-434-382-2 x CELE04A4/rev ..

Align seg 1/1 to reverse of: CELE04A4 from: 1 to: 22846

55 ProAsnThrValTyrLeuGlnValValAlaAlaGlySerArgAspSerGI 71
||||| ||| : : : : : ||| : : : : : ||| : : : : :
16210 CCAGTCAGTATCGATCGAAGTCTCGGTAATGGAAACAGGACTTCCTCG 16161
71 yAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysG 88
||||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
16160 AGCATGCTTCATCCTTCGAACACCACCTGAAACCTACATGTTCAACTGCC 16111
88 lyGluGlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArg 104
||||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
16110 CTGAGAATGCATGCACATTCCTGTGCAGCTTAGGATCCCAAGTTCACGC 16061
105 LeuAspAsnIlePheLeuThrArgMetHisTyrPheSerAsnValGlyClyLe 121
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
16060 GTCTGTGATCTTTTCATCATCATCTGCAATTCGGATAATATTGCTGGAAT 16011
121 uSerGlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValI 138
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
16010 TTCATCAATTCTACTTCT...AAGAAATCAAAATGCATTGCTACTCGAC 15964
138 euSerGlyProProLeuLeuGluLysTyrLeuAlaIleLysIlePhe 154
||| ||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
15963 TTCATGGTGCATGAATATCAAAACACTTTTGGAAATGATTTCGACCATTT 15914
155 SerGlyProLeuLysGly.....IleGluLeuAlaValAr 166
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15913 CAAGATTCGGATTACGGTAGCTGCAAGTATCCATCTCAAGTTGAAGAACG 15864
166 gProHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnI 183
||||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
15863 TCATATCAATGGAGAATTATGAGGATCCCGGATTAAAGGTACATATA 15814
183 leProIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGlnSer 199
||||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
15813 TTCCCTCG.....AGT 15803
200 ProGluArgProLeuSerArgLeuSerProGluArgSerSerAspSerGI 216
||| ||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
15802 CCA.....CCGTTGAAT.....ATCGGAAGTAAACAGA 15774
216 uSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnArgArg 233
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15773 AAAAAGTAAAAAT..... 15761
233 lyValArgAspSerSerLeuValValAlaPheIleCysLysLeuHisLeu 249
||||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
15760 ..GTAAAAGTCAACAATGTAGACATTCGATTTTAAATCGAAATGAAGGAG 15713
250 LysArgGlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProVa 266
: : : : : : : : : : : ||| : : : : : ||| : : : : :
15712 GCAGCCAGCAAGTACGATACCAAGTGTGATGGAACCTGAAGTACCGAA 15663
266 lcGlyThrAlaAlaIleAlaProIleIleAlaAlaValLysAspGlyLys 283
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15662 AGGT.....CCACTAATTGGAAACTGAAATCAGGAGAG 15628
283 erIleThr...HisGluGlyArgGluIleLeuAlaGluLeuCysThr 298
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15627 CAGTTACTCTCCAGATGCAAGGACGATTCAACCTGTATCAAGTTTCTCA 15578
299 Pro.....ProAspProGlyAlaAlaPheValValValGluCysPr 312
15577 TCGGATAAAGTTGAAGGAGACAAGCCACTGCTTTTAGTCAGAGAATGTAC 15528
312 oAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheGlnArgT 329
||||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
14639 AAGGAACATATGGCAAAATGAGACAGAGTGTGGGAGAAGTGGTTGAAG 14590

15527 TACAGAGGATCATGTGAAAGCACTGATAGACTCATCATCACTGCAACCAT 15478
329 yrGlnGly.LysAlaAspAlaProValAla..... 338
: : : : : : : : : : : ||| : : : : : ||| : : : : :
15477 TTTTGAATGGGAAAAAACAGTTGGATTATGTAAAGATTGAAAAAGAATTA 15428
339LeuValValHisMetAlaProAlaSerV 348
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15427 TTGACATTTAAAAAATACATTTTCAGATGGTTTCATATCACTGATGATGCAG 15378
348 alLeuValAspSerArgTyrGlnGlnThrMetGluArgPhe...GlyPro 363
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15377 TATCAATACTCAACCTACAGACTTTGTGGAGAAGCTCAACAATCCC 15328
364 AspThrGlnHisLeuValLeuAsnGluAsn.....Cy 374
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15327 AGTATCACTCATTTGCTGTGATCAATGGAGGAATCCTGTGATCCACCTGT 15278
374 salaSerValHis.....A 379
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15277 GGAAGTGTCTATAAGTAAGGAAAAATATGAAATATTGAATAGATCTACGA 15228
379 snLeuArgSerHisLysIleGlnThrGlnLeuAsnLeu..... 391
||| ||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
15227 ATTCGAGATCTCGGAAACAAGATCCAGGGTCTCAAGAGCTGCAGTTGCG 15178
392IleHisProAspIlePheProLeuLeuThrSer... 402
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15177 ACGATTTTCAGGACCTCTGTTTTCGAGATCTTCTATTTCAGGAGATCCCG 15128
403 .PheArgCysLysLysGluGlyProThrLeu...SerValProMetValG 418
||||| : : : : : : : : : : : ||| : : : : : ||| : : : : :
15127 TTTTCGTGTTCCTCGCGGCATCAGAAAAATTCAAAAAATTCAGTTTAC 15078
418 lngGlyGluCysLeuLysTyrGlnLeu.....ArgProArgArg 431
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
15077 AAAATCCATAGTTTTAGAAATCCAAATTTGTTATTACAGACACACTCGC 15028
432G1 432
15027 CTCTTCGATCAATTCCTCATCTCTTCCAGCTCTTCCCAATTTGA 14978
432 urp.....GlnArgAspA 437
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14977 TTGGAGTGGAAATTATACTCAAATAGGAGCTATCAACACGACAGGATC 14928
437 laIleIleThrCysAsnPro..... 443
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14927 AATTCATTCGAGTGCCTCAATGCAACGATATTGGATGCGCGAGAGCT 14878
444GluGluPheIleValGluAlaLeuGlnLeuProAsnPheG1 457
||||| ||| : : : : : ||| : : : : : ||| : : : : :
14877 TCGTTCAACGAGGACCCATCGTCAACAATTTGCTAGCTGCTGACCCAGA 14828
457 nlnSer.....ValGlnGluTyrArgArgSerAlaG 468
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14827 GCTGTCGGATAAAGCGAAGAAATTGATTAAAGAAATATCAGAAGTTGGAGA 14778
468 lnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGluIleIle 484
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14777 AGGAGAATAAATGGACTGCCGAA.....TTTCCGAAATTAACA 14740
485 PheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSerAl 501
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14739 TTTTCGGAATTCGTCGCGGTTCCTCAAAATATCGAAATGTGACAGG 14690
501 athrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyG 518
: : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14689 ATATCTGTGGAAAGCATCAGAGACAGTCCGATTTTAATTGATGTTGGAG 14640
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGly...AspGlnValAsp 533
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
14639 AAGGAACATATGGCAAAATGAGACAGAGTGTGGGAGAAGTGGTTGAAG 14590

534 ArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAs 550
 14589 CAGCTTCTGTGAATATCAACTCGCTCTTATCATCTCACCTCATCAGGA 14540
 550 pHisHisThrGlyLeuProSerIleLeuLeuArgGluArgAlaLeuA 567
 14539 TCATATGAATGGCTGTACACAATATTGCACGAGAAAGAGCTTTTCG 14490
 567 laSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGln 583
 14489 AGTCCCTCGAGCCCTTACCGTCCACTAGTCTTGATGCAATCGGAAT 14440
 584 LeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnValLeuHi 600
 14439 GTTCTGAACCAATGAAGACCTAC...TCAATTTGCTTTGAAACATTTGA 14393
 600 sHisIleSerMetIle..... 605
 14392 GCATCTCTCGAATGTTGATATCTCCAGGTATCCACTAACACCTCCCG 14343
 606ProAlaLysCysLeuGlnGluGlyAlaGluIle 616
 14342 GGAGCCCTGGAGGCCCTCTCGAAAG...ACCGAGACTTCGGAGCCC 14297
 617 SerSerProAlaValGluArg...LeuIleSerSerLeuLeuArgThrCysA 633
 14296 ACATCTCCCGCAGCAGACAGATGTTCTACAGAGATATGTCAGCAGTTTCG 14247
 633 sp.....LeuGluGluPheGlnThrCysLeuValArgHis 644
 14246 ATAAAGGATGAGAGCTGGATGCTGAAGGCTGTACAGTTTCATCAT 14197
 645 CysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLysVa 661
 14196 ACAAGATGCAATGGA...TTTGATGAGAGTTGCTGGAACCAAT 14150
 661 lValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMetGlyL 678
 14149 TGTATCTCTGGAGATACGAAACCATGATTTGTTGTTGAAGAAGAA 14100
 678 ysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly..... 692
 14099 AGATGCTGATGTTCTTGTTCATGATCCACTTTTGAAGATGGTCATGAG 14050
 692 692
 14049 GTAGATATGACACCAACCCCAAAAAAATTCGCTAAATTTGCTCTTT 14000
 692 692
 13999 GGTTTTATTACTAAATGTACGAGATTTTAGCCAAAAATACGGTAGTAG 13950
 693LeuGluGluGluAla 697
 13949 GGTCTCAACGACAAATTTTGGATAAATGCAGAGATTGAAGATAACAGA 13900
 698 Val..... 698
 13899 ATTTTCATCGAATTTTCATAGTTTTCGTTGAGAAATATATAAATCTGTG 13850
 698 698
 13849 AGAAAAATCGATGAATCTTTCAGAAACAAAGTTTGAACCTACAGTAT 13800
 698 698
 13799 TCITCAAAGCGAATAACTTTTTCATTCGGCAAAAAATTTTGTGCTGTG 13750
 698 698
 13749 AGACCCATATTACGTATTTTGTAGTGCANATGAAAAATCCAACTATTA 13700

698 698
 13699 GAAACCCGCTTTTTCAGTTTTTTTTTTTAAATTTGAATTCACCAACTG 13650
 698 698
 13649 AGAATTTCAAGACCAATCAGCGAGCTTCTCTCTCATCTCCGCTGATTTGG 13600
 698 698
 13599 TCGAATATGGCGGAGCTATTTCGCTGATTTGTCATATTTCATTTGGC 13550
 698 698
 13549 AGGCATTTCAAACACGCAATTTTTCAGTTTCGGAATTTTTCGTTT 13500
 698 698
 13499 TAAGGCAATATTGAAATTTTTCACCAATTAATTTTCAGGCTGATGCANT 13450
 699 GlulysThr.HisSerThrThrSerGlnAlaIleSerValGlyMetArgM 715
 13449 GAGAAACGTCATTTCAACAATGGGTCAAGCAGTCGATGTTGAAACGAA 13400
 715 etAsnAlaGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLys 731
 13399 TGAATGCCAACACATATTTCGACATTTTCAGTTCGAGATATCCAAA 13350
 732 ValProLeuPheSerProAsnPhe.....SerGluLysValGlyValAl 746
 13349 GTACCAATCTT...CTGAATATCTTGATTAAGGAANAATATTGGTGTGC 13303
 746 aPheAspHisMetLysValCysPheGlyAspPheProThrMetProLysL 763
 13302 GATGGATGTTGAGAGTTCGATTCGATCATTTACCACCTTCTTTCGAAC 13253
 763 euIlePro 765
 13252 TTTCTCC 13245

seq_name: gb_htgl:AC006719

seq_documentation_block:

LOCUS AC006719 209365 bp DNA HTG 23-FEB-1999
 DEFINITION Caenorhabditis elegans clone Y17G9, *** SEQUENCING IN PROGRESS ***,
 2 unordered pieces.

ACCESSION AC006719
 VERSION AC006719.1 GI:4263219
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 209365)
 Waterston, R.H.
 The sequence of *Caenorhabditis elegans* clone
 Unpublished
 2 (bases 1 to 209365)
 Waterston, R.H.
 Direct Submission

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 3306: contig of 3306 bp in length
 * 3307 3319: gap of unknown length

MO 63108, USA

Submitted (23-FEB-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

25316 AGGAGATAAATGGACTCGAA.....TTTCGGAATTAACA 25279
 485 pheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnValSerAl 501
 25278 TTCTTCGGAACCTCGTCCGCGTTCATCAAAATATCGAATGTGACAG 25229
 501 aThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyG 518
 25228 ATATCTGGTGGAGCATCAGACACAGTCCGATTTAAATTGATGTGGAG 25179
 518 luGlyThrPheGlyGlnLeuCysArgHisTyroGly...AspGlnValAsp 533
 25178 AAGGAACATATGCACAATAGAGAGCAGTTTGTGAGAAGATGTTGTAAG 25129
 534 ArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAs 550
 25128 CAGCTTCTGTGTAATCACTCACTCGCTTATCATCTCAGCTCATCAGGA 25079
 550 pHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuA 567
 25078 TCATATGATGGCTGTACACAATTTATGCACGGAGAAAGAGCTTTTCG 25029
 567 leSerLeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGln 583
 25028 AGTCCCTCGAGGCCCTTACCGTCCACTAGTCTGTATGCAATCGGAAT 24979
 584 LeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHi 600
 24978 GTTCTGAAACCAATGAAGACCTAC...TCAATTTGCTTTGAAACCATTTGA 24932
 600 sHisIleSerMetIle..... 605
 24931 GCATCTCTCGAAATTTGTTGATATCTCCAGGTATCCACTAACACCTCCCG 24882
 606ProAlaLysCysLeuGlnGluGlyAlaGluIle 616
 24881 GGAGCCCTGGAGGCCCTCTCGGAAAG....ACCAGACTTCCGAGGCC 24836
 617 SerSerProAlaValGluArg...LeuIleSerSerLeuLeuArgThrCysA 633
 24835 ACATCTCCGCGCCGAGCAGAGATGTTCTACAGGATATGTCAGCAGTTTCG 24786
 633 sp.....LeuGluGluPheGlnThrCysLeuValArgHis 644
 24785 ATAAAAGGCATGGAAGCTGGATGAGCTGAAGGCTGACAAAGTTCATCAT 24736
 645 CysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLysVa 661
 24735 ACNAGATGGCAATGGA...TTTGTGATGAGAGTTGCTGGAAACGCAAT 24689
 661 lValTySerGlyAspThrMetProCysGluAlaLeuValArgMetGlyL 678
 24688 TGTATCTCTGGAGATACGAAACCATGTGATTTGTTGGTTGAAGAAGAA 24639
 678 yAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly..... 692
 24638 AAGATGCTGATGTCTGTGTTTCATGAATCCACTTTTGAAGATGTCATGAG 24589
 692 692
 24588 GTAGATATGACACCAAAACCCCAAAAATTTGGCTAAAAATTTTCGTCTTT 24539
 692 692
 24538 GGTTTATTACTAAATGTACGAGATTTTATAGCCAAAATACGGTAGTAG 24489
 693LeuGluGluGluAla 697
 24488 GGTCTCAACACGACAAATTTTGTGATAAATGACGAGATTTGAAGATAACAGA 24439
 698 Val..... 698
 24438 ATTTTCATCGAATTTTCATAGTTTTTTCGTTGAGAAATATATAAATCTGTG 24389

698 698
 24388 AGAAAAATCGATGAAATCTTGAGAAACAAAAAGTTTGAACCTACAGTAT 24339
 698 698
 24338 TCTTCAAGCGCAATAACATTTTTCGATTCGGCAAAAATTTTGTGCTGTCG 24289
 698 698
 24288 AGACCTATATATTACGTATTTTGTAGTCAATAATGAAAAATCCAACATTATTA 24239
 698 698
 24238 GAAACCCGCTTTTTCAGTTTTTTTTTTTAAATTTGAATTCATCCAAACTG 24189
 698 698
 24188 AGAATTCAGACCAATCAGGACTTGTCTCTCATTCGCGCTGATTTGG 24139
 698 698
 24138 TCGAATATGCGCGGAGCTATTTCGCTGATTGTGTATTTCTCATTTGGC 24089
 698 698
 24088 AGCATTTCAAACAGCAATTTTTCAGTTTCGGACTATTTTGGTTT 24039
 698 698
 24038 TAAGCAAAATATTGAAATTTTCAACCAATTAATTTTCAGGCTGATGCAAT 23989
 699 GluLysThr.HisserThrThrSerGlnAlaIleSerValGlyMetArgM 715
 23988 GAGAAACGTCATTCACAAATGGGTCAAGCAGTCGATGTTGGAAACGAA 23939
 715 etAsnAlaGluPheIleMetLeuAsnHisPheSerGlnArgTyrAlaLys 731
 23938 TGAATGCCAAGCACATATTCTGCACACATTTCAGTCGAGATATCCAAAA 23889
 732 ValProLeuPheSerProAsnPhe.....SerGluLysValGlyValAl 746
 23888 GTACCAATTTCTT...CCTGAATATCTTGATAAGAAATATTGTGTGGC 23842
 746 aPheAspHisMetLysValCysPheGlyAspPheProThrMetProLysL 763
 23841 GATGGATATGTTGAGAGTTCGATTCGATTCATTTTACCACCTGTTTCGAAC 23792
 763 eullePro 765
 23791 TTCTTCCC 23784
 seq_name: gb_p12:AP001550
 seq_documentation_block:
 LOCUS AP001550 143209 bp DNA PLN 29-MAR-2000
 DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0431F01.
 ACCESSION AP001550
 VERSION AP001550.1 GI:7340902
 KEYWORDS
 SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0431F01.
 ORGANISM
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
 REFERENCE 1 (bases 1 to 143209)
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
 TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone:P0431F01
 JOURNAL Published Only in Database (2000) In press
 REFERENCE 2 (bases 1 to 143209)
 AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
 TITLE Direct Submission

JOURNAL

Submitted (27-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@abr.affrc.go.jp, URL:http://www.dna.affrc.go.jp:82/, Tel:81-298-38-7441, Fax:81-298-38-7468)

COMMENT

The orientation of the sequence is from SP6 to T7 of the PAC clone. Genes were predicted from the integrated results of the following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October1998 version). The genomic sequence was searched against the non-redundant database NRP(PIR,SWISSPROT, GENPEPT, PDB) from MAFF DNABank and the cDNA sequence database at RGP. Protein similarities of the coding regions were searched against NRP with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN2.0 with the corresponding DBJ accession no. and RGP clone ID. This sequence of P0431F01 clone has an overlap with P0485D09 clone at the 5' end. The sequence of this clone starts at the position 93,219 of P0485D09. Detailed information on assemble quality together with annotation of this entry at <http://www.dna.affrc.go.jp:82/genomicdata/GenomeFiniished.html>.

FEATURES

source

Location/Qualifiers

1..143209
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0431F01"
complement(join(10340..14224,14755..14802))
/note="Similar to AF049110.1(AF049110|pid:94206306) Zea mays retrotransposon Cinfu1-1; internal stop codons:13700..13702,13391..13393,13272..13274, 13086..13088,13020..13022,11549..11551,11558..11560, 10880..10882; probably inactive because stop codons and frameshift positions are included in CDS"
/pseudo

misc_feature

10917..10918
/note="one bp frameshift deletion"
11429..11430
/note="one bp frameshift deletion"
12388..12389
/note="one bp frameshift deletion"
12743..12744
/note="one bp frameshift deletion"
13290..13291
/note="two bp frameshift insertion"
complement(join(17293..17493,18173..18283,18920..18973))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92975.1"
/db_xref="GI:7340903"

CDS

/translation="MPTVQIIPRTGCMVFILCNFDLIASSVCVCANMOCIMAGERST DGLPFPAGQAEKDGKQKQAKDDLSRGCLVPSVCMPLHADNDVVGSDFWA AAGGGGGAPPLAGNNLR"
complement(join(20649..20849,20887..21012,21025..21150, 21462..21521,23357..23422,24370..24668,25123..25368, 28863..29059,29676..29733,31244..31316,31424..31531))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92976.1"
/db_xref="GI:7340904"

CDS

/translation="MNRNKGTPRGSGARPGGVGVRGVRAMAAAAAARWRPRL PGAGDEAGDHMRGPNQNDIISSRLSRPKKRLPSLHVNNVAIPELGRPDMRR RGSVDMGRKSGALSCFYLCSESPRIELTYTNKRTSMLSGGELKSSSL VQMVVGTGNSSSIMGSLRQPCSEQDAASSPAMFLPQQLLLHASNSPL NIPENLSTGLDDASMPHEASAAAYDFYGHGGAGDGILOASPEASSCKSLSOMLL QAAASPRCVSTTSGGSSMDFSNATAVAAPAEPELTKRHAGQSDNSSEVKERLG DRTALHQLIVSPGKALSYPLGQCCSNPMQOOTVKPQNPQPHTRWPVWCMSAAHD ATGACACAAAHRRGHSILRASCWSLAEAVQPLVVGHRRADELSPQQLVPL PGRVACAAQVLRASHRCCSSLAARAPAAATCCRAPPTRRRALARAFAHRRRTATP

CDS

DPRLRTRCRA"
join(33172..33273,33899..34024)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92977.1"
/db_xref="GI:7340905"
/translation="MTIFRYCYCEKVVMLHDDCYVTAKFLRDVNNNOVISISTIGDGR PFNTCKIRLIMFISRPVHRSLAGHPQ"
complement(join(34876..35013,36635..36695,36723..36861, 37054..37169,38057..38157))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92978.1"
/db_xref="GI:7340906"
/translation="MHTIRCISAGALVLMVNLNKEELYVVGSSNOFFNSOHLSP HLYLLSPHPPSLSNLPDVAAGGHTVAVLHPAAGGGRSTAEVARGVWRK EAAAAIIPVDLAADGFFSRADDISTRTRKTFDSVRMDGISSTGSEQLAAPPSS APASELGSHPSSRRRRRHDPGTPA"
join(40151..40508,41020..41387)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92979.1"
/db_xref="GI:7340907"
/translation="MGTEKKHRQSSRRSRKRDVTGKGRCEVASKASDGRVQKRHH KQGYTRTLPSPFLXVLDRTLOPDASLAAAFHPCLPACTRSRSSELARCCAC FGKGVQGEYERCHSVPGGLEGIDTNPTRTLAMVGTSSGNLSNGDDNYGSSIFS LKSPLAWTOPPSLRTGSSTMRRCDGLDLPDGPDTTDRSTATIGVWRGSRGGGARR GNSLDDDELGLGDGATADER"

CDS

join(42260..42269,42373..42761)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92980.1"
/db_xref="GI:7340908"
/translation="MCTSGAGKSSFOASDPDSARRVLTITCRPAGESIDAAPCCTTR DVPRAGYERPSPAHEHTLSLKSASYGKRFEMAYVLKLSLTAVCTVGNVNRKV WGVPPAKTDLAKMLQFLAAHFKTPLAKI"
complement(join(46089..46586,47173..47257,48631..48762, 49150..49229))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92981.1"
/db_xref="GI:7340909"
/translation="MNNAKFFYNNMKRKYKISPRNDGVQDRSCLIKDLFGGOLLETPP SASIKVNLRLDVTYHNSQMVYQIRSSNRDDTPRIYQKFGSDRSTSVSNCDQRPD ITHHELSPDELPQSRAYSVAYAWOPSOHFLKDGAAHLSVTLSSLPLALS LSLSRGHSAATGRRRRAGHGGVGGVGRAGGGRGGGAGGCGVCGDGPSCN PRDGRSLPRPPAMPCCATTALANPPYRGALAREKEEQR"

CDS

join(50116..50125,50292..50697,51472..51704,52558..52654, 52775..52836,53717..53754)
/note="ESTs AU083004(E60493),AU030968(E60493), D23864(R0457) correspond to a region of the predicted gene.; Similar to Glycine max GH1 protein (AF016633)"
/codon_start=1
/protein_id="BAA92982.1"
/db_xref="GI:7340910"
/translation="MYTGITRPIRVHVLPLPMSPPLELDYIGLSPPPPPPSSSSAAA ARDDYDLKGTLELRLPGSESPDRPAAATAAAATATLLELPAGKAKVFPDEALN TPPTAAGKGAAREGEVECAEDKVAAPPOPAKAOVVGHPIRSYKNTMATNO IKSNEVDVAKQCGELYVKVSMGDPAPLKKVLDKTYKNYKMSLGLKMFIFGFTGK EAGENQDGEIVLITYEDKDDMWLVGDVPWEMFTDCRRRLRMKGSDAIGLAPRAGEK SKNRN"

CDS

complement(join(54434..54639,56954..57117,57172..57379, 57474..57483))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92983.1"
/db_xref="GI:7340911"
/translation="MRDLSLCOEKKKKRREAPAROLGSGTVECVQTVRRFA ROVSGSSHVGTAGAHVQNPSPRTRCTPPIFALTSGPRTVVVSVLGMCTCSWG ADVDRSSVGPVSDAVGPRVRGALHQAHLQADTNSNGKEKEKERTENSIGGK QKOQILQARGGKESLILRDHPSLFTLHLYFYTT

CDS

join(60274..60854,60878..60896,61212..61688)
/note="hypothetical protein"

CDS

779 uGluArqArqGlu 783

OM of: US-09-434-382-2 to: N_Geneseq_36:* out_format : pfs
Date: Feb 18, 2001 9:25 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=framet-p2n.model -DEV=xlpl
-Q/cgnl_1/USPTO.spool/US09434382/runat_16022001_105749_29718/app_query.fasta_1.895
-DB=N_Geneseq_36 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=200000000
-USPR=US09434382_CGnl_1_0 -NCPU=6 -LCPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-434-382-2
Query length: 826
Database: N_Geneseq_36:*
Database sequences: 480022
Database length: 187831343
Search time (sec): 105.780000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgnl_8/gcgdata/geneseq/NA2000.DAT:280231 +			404.00	613.88	238
/cgnl_8/gcgdata/geneseq/NA1998.DAT:296286 -			221.50	305.90	3258
/cgnl_8/gcgdata/geneseq/NA1998.DAT:52183 -			217.50	294.27	5338
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X20534 +			195.00	242.57	25187
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X13218 +			187.50	247.47	5830
/cgnl_8/gcgdata/geneseq/NA2000.DAT:A26852 +			175.50	238.63	2494
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X22740 +			172.50	168.09	910715
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X22028 +			131.50	162.27	5660
/cgnl_8/gcgdata/geneseq/NA1997.DAT:T58899 -			129.00	166.76	2692
/cgnl_8/gcgdata/geneseq/NA1997.DAT:X30827 +			124.50	177.62	551
/cgnl_8/gcgdata/geneseq/NA1998.DAT:T98612 +			123.50	165.46	1426
/cgnl_8/gcgdata/geneseq/NA1998.DAT:X34766 +			123.00	143.20	9759
/cgnl_8/gcgdata/geneseq/NA1995.DAT:X90529 +			122.00	153.13	3501
/cgnl_8/gcgdata/geneseq/NA1988.DAT:N80456 +			121.50	154.14	2987
/cgnl_8/gcgdata/geneseq/NA1997.DAT:T59895 +			120.50	150.84	3501
/cgnl_8/gcgdata/geneseq/NA1997.DAT:T89842 +			120.00	138.62	9759
/cgnl_8/gcgdata/geneseq/NA2000.DAT:A08043 +			120.00	138.62	80161
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X21501 +			119.00	113.55	19024
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X13116 +			118.50	128.87	50341
/cgnl_8/gcgdata/geneseq/NA1998.DAT:X22674 +			116.50	114.93	50341
/cgnl_8/gcgdata/geneseq/NA2000.DAT:X39519 +			116.50	114.93	52297
/cgnl_8/gcgdata/geneseq/NA1995.DAT:T51411 +			116.50	114.51	52298
/cgnl_8/gcgdata/geneseq/NA1993.DAT:Q47357 +			115.50	131.74	9762
/cgnl_8/gcgdata/geneseq/NA2000.DAT:X48249 +			115.00	139.53	4541
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X72100 +			115.00	131.74	29879
/cgnl_8/gcgdata/geneseq/NA1993.DAT:Q46806 +			115.00	118.47	9757
/cgnl_8/gcgdata/geneseq/NA1995.DAT:Q97686 +			114.50	130.22	2339
/cgnl_8/gcgdata/geneseq/NA1996.DAT:T35251 -			114.00	145.42	3109
/cgnl_8/gcgdata/geneseq/NA1998.DAT:X34853 +			114.00	142.24	6400
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X22707 +			114.00	134.17	2359
/cgnl_8/gcgdata/geneseq/NA1996.DAT:T31273 +			113.50	144.56	5510
/cgnl_8/gcgdata/geneseq/NA2000.DAT:X51671 +			113.50	135.08	6847
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X98922 +			113.50	132.65	42235
/cgnl_8/gcgdata/geneseq/NA1999.DAT:X98035 +			113.50	112.31	44377
/cgnl_8/gcgdata/geneseq/NA1997.DAT:T78508 +			113.50	111.76	44377
/cgnl_8/gcgdata/geneseq/NA1997.DAT:T80414 +			113.50	111.76	77536
/cgnl_8/gcgdata/geneseq/NA2000.DAT:A14651 +			113.50	105.52	5970
/cgnl_8/gcgdata/geneseq/NA2000.DAT:X256003 +			113.00	133.42	1922
/cgnl_8/gcgdata/geneseq/NA1994.DAT:Q66639 +			112.00	144.56	3279
/cgnl_8/gcgdata/geneseq/NA1993.DAT:Q49397 +			112.00	138.59	

/cgnl_8/gcgdata/geneseq/NA2000.DAT:000334 - 111.50 139.43 5.61 2842
/cgnl_8/gcgdata/geneseq/NA1995.DAT:087426 + 111.00 141.51 4.29 2204
/cgnl_8/gcgdata/geneseq/NA1998.DAT:X30458 - 110.50 79.36 1.2e+04 53472
/cgnl_8/gcgdata/geneseq/NA1998.DAT:X30459 - 110.50 79.32 1.2e+04 53616
/cgnl_8/gcgdata/geneseq/NA1990.DAT:Q06470 - 110.00 136.22 8.46 3086
seq_name: /cgnl_8/gcgdata/geneseq/NA2000.DAT:280231
seq_documentation_block:
ID 280231 standard; CDNA; 238 BP.
AC 280231;
XX 07-APR-2000 (first entry)
XX Human colon cancer cell line SW480 cDNA clone SEQ ID NO:315.
XX Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia;
KW hyperplasia; ds.
XX Homo sapiens.
OS
XX WO9964576-A2.
PW
XX 16-DEC-1999.
PD
XX 09-JUN-1999; 99WO-IB01062.
PF
XX 10-JUN-1998; 98US-0088801.
PR
XX (FARB) BAYER CORP.
PA
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JE;
PI Schlegel R;
XX WPI; 2000-087220/07.
DR
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer -
XX Claim 15; Page 258; 469pp; English.
CC Z79917 to 280766 represent double stranded cDNA clones isolated from the
CC human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which
CC can be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia.
XX Sequence 238 BP; 55 A; 57 C; 69 G; 57 T; 0 other;
SQ

alignment_scores:
Quality: 404.00 Length: 78
Ratio: 5.179 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-434-382-2 x 280231 ..
Align seg 1/1 to: 280231 from: 1 to: 238
60 LeuGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrrVa 76
|||||
3 CTGACGGTGTGGCGGGTACGGGGAGTCGGGGCGCGCTCTACGT 52

76 lPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGlyValGlnA 93
 |||||
 53 CTTCTCCAGATTCAACCGGTATCTTCAACTGTGGAGAAGCGCTTCAGA 102
 |||||
 93 rGLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIlePhe 109
 |||||
 103 GACTCATGACAGAGACAAAGTTAAAGTTGCTCGCTGGACAACATATTC 152
 |||||
 110 LeuThrArgMetHisTspSerAsnValGlyGlyLeuSerGlyMetIleLe 126
 |||||
 153 CTGACACGAATGCAGCTGCTCTAATGTTGGGGCTTAAGTGAATGATTCT 202
 |||||
 126 uThrLeuLysGluThrGlyLeuProLysCysVal 137
 |||||
 203 TACTTTAAAGGAACCGGGCTTCCAAGTGTCTA 236

seq_name: /cgn1_8/gcgdata/geneseq/geneseq/NAL1998.DAT:Z96286

seq_documentation_block:

ID Z96286 standard; DNA; 3258 BP.

XX Z96286;

XX AC

DT 10-APR-2000 (first entry)

XX S. pneumoniae derived DNA from ORF #114.

XX Treatment; prevention; disease; diagnosis; gene therapy; screening;
 KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.

XX Streptococcus pneumoniae.

XX WO9806734-A1.

XX 19-FEB-1998.

XX 15-AUG-1997; 97WO-0514436.

XX 16-AUG-1996; 96US-0024022.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
 PI Stodola RK;

XX WPI; 1998-159452/14.

XX P-PSDB; Y85922, Y85923, Y85924.

XX Streptococcus pneumoniae proteins and related DNA - useful for
 PT screening compounds for antibacterial activity

XX Claim 4; Page 141-142; 640pp; English.

XX This invention describes novel isolated Streptococcus pneumoniae
 CC polynucleotides (see Z96173-Z96494) and their encoded proteins (see
 CC Y85792-Y86182). The DNA, vectors and host cells described in the method
 CC of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX Sequence 3258 BP; 907 A; 723 C; 675 G; 953 T; 0 other;

alignment_scores:

Quality: 221.50 Length: 332
 Ratio: 1.393 Gaps: 10
 Percent Similarity: 47.892 Percent Identity: 24.398
 alignment_block:
 US-09-434-382-2 x Z96286/rev ..
 Align seg 1/1 to reverse of: Z96286 from: 1 to: 3258
 482 GluIlePheLeuGlyThrGlySerAlaIlePrometLysIleArgAs 498
 : : : : : |||||
 1851 GATATTCAATTTTAGGAACGGGGCTGGTCAGCCCTTAAGCCCGCAA 1802
 : : : : : |||||
 498 nValSerAlaThrLeuValAsnIleSerProAspThrSer.....LeuL 513
 |||||
 1801 CGTTTCAAGTCTCGCCCTGAAACTTTTGGACGAGATTAAACGAAGTTTGGC 1752
 : : : : : |||||
 513 euLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGly 529
 |||||
 1751 TCTTTGACTGTGGAGAGGTAGC..... 1729
 : : : : : |||||
 530 AspGlnValAspArgValLeuGlyThr.....LeuAl 540
 : : : : : |||||
 1728CAAAATCGCATCTCTGAAACCAACAATTCGACCACGTAAGTGCAG 1685
 : : : : : |||||
 540 aAlaValPheValSerHisLeuHisAlaAspHisHisThrGlyLeuProS 557
 : : : : : |||||
 1684 CAAATCTTTATCCCATCTGTCATGGAGACCACATTTTGGTTGCCAG 1635
 : : : : : |||||
 557 erIleLeuLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeu 573
 : : : : : |||||
 1634 GTTTCCTTTCTAGCCGT.....GCCTTTAGGCCCAATGAAGAGCAG 1594
 : : : : : |||||
 574 HisProLeuLeuValAlaProAsnGlnLeuLysAlaTrpLeu..... 588
 : : : : : |||||
 1593 ACAGATTGGAAATCTACGGACCTCAGGAATCAAGTCATTGTCTTAAC 1544
 : : : : : |||||
 588 588
 1543 CAGCCTTCGTGTGTGTCAGTTCTCGTCTGCCCTACCGCATTCATTTCATG 1494
 : : : : : |||||
 589GlnGlnTyrHis 592
 : : : : : |||||
 1493 AGTTTGACCAAGATTCTCTGGTAAAATTTTGAACCGATAAAATTCAC 1444
 : : : : : |||||
 593 AsnGlnCysGlnGluValLeuHisHisIleSerMetIleProAlaLysCy 609
 : : : : : |||||
 1443 GTGTATGCAGAGAGCTGGACCACACTATTTTCTGTGTGGCTATCGTGT 1394
 : : : : : |||||
 609 sLeuGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleS 626
 : : : : : |||||
 1393 CATGCAAAAGCATCTAGAA..... 1375
 : : : : : |||||
 626 erSerLeuLeuArgThrCysAspLeuGluPheGlnThrCys..... 640
 : : : : : |||||
 1374GGGACCGTGGATGCTGAAAAACCTCAGGCTGCTGGTGT 1336
 : : : : : |||||
 641LeuValArgHisCysLysHisAlaPheGlyCysAlaLe 653
 : : : : : |||||
 1335 CCGTTCGGCCGCTTTTGGTAAATCAANAAC.....GGCCAGGATCT 1292
 : : : : : |||||
 653 uValHisThrSerGlyTrpLys..... 660
 : : : : : |||||
 1291 TGTTTTGGAGACGCACTGAAATCAAGGCAGACACTATATCTCAGCGC 1242
 : : : : : |||||
 661ValValTyrSerGlyAspThrMetProCysGlu 671
 : : : : : |||||
 1241 CACGTCCAGGTAAAGATTATCATTATTTAGGAGACACTCGAAAAACGGAT 1192
 : : : : : |||||
 672 AlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaTh 688
 : : : : : |||||
 1191 GCCAGTGTGCTGTGGCTGTCAATGCAGATGCTCTAGTTATGATGTCAC 1142
 : : : : : |||||

```

688 rLeuGluAspClyLeuGluGluAlaValGluLysThrHisSerThr 705
1141 TTATGGAAGGTGATGAAAAATTTGCTCGTAACCATGGTCACTCAACTA 1092
705 hrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMet 721
1091 ATATGCAAGCTGCACAGTAGTACCGGTAGAACGAGGTGCCAAACGCTCTTA 1042
722 LeuAsnHisPheSerGlnArgTyr.....AlaLysValPr 733
1041 CTCACCATATCAGTGCCTCGTTCCTCTCAAAAGATATAGCAAACTCAA 992
733 oLeuPheSerProAsnPheSerGluLysValGlyValAlaPheAsp 748
991 GAAGACGCTGCCCAATTTTGAATAATGTCCATGTGTCAAAGAC 946

```

seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA1998.DAT:V52183

seq_documentation_block:

ID V52183 standard; DNA; 5338 BP.

XX AC V52183;

XX DT DT

XX DE 23-OCT-1998 (first entry)

XX DE Streptococcus pneumoniae genome fragment SEQ ID NO:50.

XX KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay; computer readable medium; vaccine; pharmaceutical composition; ds.

XX OS Streptococcus pneumoniae.

XX PN WO9818931-A2.

XX PD 07-MAY-1998.

XX PF 30-OCT-1997; 97WO-US19588.

XX PR 31-OCT-1996; 96US-0029960.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

XX PI Kunsch CA, Rosen CA;

XX DR WPI; 1998-272225/24.

XX PT Computer-readable medium with recorded Streptococcus pneumoniae polynucleotide sequences - useful in diagnostic kits and assays, and pharmaceutical compositions and vaccines for Streptococcus pneumoniae

XX PS Claim 1; Page 452-455; 1409pp; English.

XX CC The present invention describes a computer readable medium which has the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded on it, or a representative fragment or a sequence at least 95% identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1 to 391 (V52134 to V52524) are genomic fragments from Streptococcus pneumoniae. The present invention also describes an isolated nucleic acid molecule encoding a homologue of any of the fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced by a process comprising: (a) screening a genomic DNA library using as a probe a target sequence defined by any of the sequences in SEQ ID NO:1 to 391, identifying members of the library which contain sequences that hybridize to the target sequence and isolating the nucleic acid molecules from the members; or (b) isolating mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid molecules whose nucleotide sequence is homologous to amplification primers derived from the fragment of the S. pneumoniae genome to prime the amplification and isolating the amplified sequences. The computer readable medium can be used in a computer-based system for identifying fragments of the S. pneumoniae genome of commercial importance, or expression modulating

CC fragments of the S. pneumoniae genome. Products from the present invention can be used in diagnosis kits and assays, and pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 5338 BP; 1476 A; 1274 C; 1092 G; 1496 T; 0 other;

alignment_scores:

Quality: 217.50 Length: 332
Ratio: 1.377 Gaps: 10
Percent Similarity: 47.590 Percent Identity: 24.398

alignment_block:

US-09-434-382-2 x V52183/rev ..

Align seg 1/1 to reverse of: V52183 from: 1 to: 5338

```

482 GluIleIlePheLeuGlyThrGlySerAlaIlePrometLysIleArgAs 498
2654 GATATCAATTTTAGAACGGGGCTGGTCAGCCCTCTAAAGCCCGCAA 2605
498 nValSerAlaThrLeuValAsnIleSerProAspThrSer.....LeuL 513
2604 GCTTCAAGTCTCGCCCTGAAACTTTTGGATGAGATTAAACGAAGTTGGC 2555
513 euLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGly 529
2554 TCTTTGACTGTGGAGAGGTACG..... 2532
530 AspGlnValAspArgValLeuGlyThr.....LeuAl 540
2531 .....CAAAATCGCATTTCTGGAACCAACAATTCGACCACGTAAGTCCAG 2488
540 aAlaValPheValSerHisLeuHisAlaAspHisHisThrGlyLeuProS 557
2487 CAAATCTTTATTACCCATCTGCATGAGACACACATTTTGGTTTGGCCAG 2438
557 erIleLeuLeuGlnArgAlaLeuAlaSerLeuGlyLysProLeu 573
2437 GTTTCCTTTCTAGCGT.....GCCTTCAGGCCAATGAAGACGAG 2397
574 HisProLeuValValAlaProAsnGlnLeuLysAlaTyrLeu..... 588
2396 ACAGATTGGAAATCTACGGACCTCAAGGAATCAAGTCATTGTCTTAAC 2347
588 ..... 588
2346 CAGCCTTCGTGTGCAGGTTCTCTCTCCCTACCGCATTCATTTTCCATG 2297
589 .....GlnGlnTyrHis 592
2296 AGTTTGACCAAGATTCTCTGGTAAATTTCTTGAACCGATAAATTCAC 2247
593 AsnGlnCysGlnGluValLeuHisHisIleSerMetIleProAlaLysCy 609
2246 GTGTATGCAGAGGAGCTGACACACTATTTTCTGTGTGGCTATCGTGT 2197
609 sLeuGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleS 626
2196 CATGCAAAAGATCTAGAA..... 2178
626 erSerLeuLeuArgThrCysAspLeuGluPheGlnThrCys..... 640
2177 .....GGACGCTGGATGCTGAAANAACCTCAAGGCTGTGGTGT 2139
641 .....LeuValArgHisCysLysHisAlaPheGlyCysAlaLe 653
2138 CCGTTCGGCCCGCTTTTGGTAAATCAAAAC.....GGCCAGGATCT 2095
653 uValHisThrSerGlyTyrLys..... 660
2094 TGTTTGGAAACGGAATGAATCAAGGCAGACACTATATCTCAGCGC 2045

```

```

661 .....ValValTyrSerGlyAspThrMetProCysGlu 671
2044 CACGTCCAGGTAAAGATTATCACTATTTTAGGAGACACTCGAAACGGGT 1995
672 AlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaTh 688
1994 GCCAGTGTGGCTGGCTCAATGACAGATGCTTAGTTTCATGAGTCCAC 1945
688 rLeuGluAspGlyLeuGluGluGluAlaValGluLysThrHisSerThrT 705
1944 TTATGGCAAGGTGATGAAAAAATTGCTGCTAAACCATGCTCACTCAACTA 1895
705 hrSerClnAlaIleSerValGlyMetArgMetAsnAlaGluPheIleMet 721
1894 ATATGCAAGTGCACAAGTAGCGGTAGAGCAGGTGCCAAACGCCCTCCTA 1845
722 LeuAsnHisPheSerGlnArgTyr.....AlaLysValPr 733
1844 CTCACCATATCAGTCCCGCTTCTCTCAAAAGATATTAGCAAACTCAA 1795
733 oLeuPheSerProAsnPheSerGlyLysValGlyValAlaLalaPheAsp 748
1794 GAAAGACGCTGCCACAATTTTGAATAATGTCATGTGGTCAAAAGAC 1749
seq_name: /cgnl_8/gcdata/geneseq/geneseqn/NA1999.DAT.X20534

```

seq_documentation_block:

X20534 standard; DNA; 25187 BP.

X20534;

05-MAY-1999 (first entry)

Polynucleotide sequence from the genome of *Treponema pallidum*.

Treponema pallidum infection; syphilis; Borrelia infection; animal;
enzyme production; ds.

Treponema pallidum.

W09859034-A2.

30-DEC-1998.

23-JUN-1998; 98WO-US13041.

24-JUN-1997; 97US-0050667.

(HUMA-) HUMAN GENOME SCI INC.

Fraser CM;

WPI; 1999-081273/07.

New isolated *Treponema pallidum* nucleic acids - used to develop
products for the detection, diagnosis, characterisation, prevention
and therapy of *T. pallidum* infections, particularly syphilis

Claim 1; Page 375-389; 1150pp; English.

X20500-21243 represent polynucleotide sequences from the genome of
Treponema pallidum. The sequences can be used for detection,
diagnosis, characterisation, prevention and therapy for *T. pallidum*
infections, particularly syphilis. They can also be used for detecting
diseases related to *Borrelia* infections in animals, and for the
production of biosynthetic products such as enzymes.

Sequence 25187 BP; 6051 A; 7307 C; 6131 G; 5671 T; 27 other;

alignment_scores:

Quality: 195.00
Ratio: 1.010

Length: 393
Gaps: 13

Percent Similarity: 49.109 Percent Identity: 21.628
alignment_block:
US-09-434-382-2 x X20534/rev ..
Align seg 1/1 to reverse of: X20534 from: 1 to: 25187

```

395 AspIlePheProLeuLeuThrSerPheArgCysLysLysGluGlyProTh 411
14265 GAGATCTTTCCGCTCACCACAGCTTCCGGAAGGACAGACATCGCGTACTC 14216
411 rLeuSerValProMetValGlnGlyCysLeuLeuLysTyrGlnLeuA 428
14215 AACAGCGGTGAATTTTGAACCGTCGAATCGCTATGCTGCTATCGCCTGG 14166
428 rgProArgArgGluTrpGlnArgAspAlaIleIleThrCysAsnProGlu 444
14165 GACTTCCGCGCCAGTACGACGAGAC.....AAA 14137
445 GluPheIleValGluAlaLeuGln.....LeuProAsnPh 456
14136 GGCTTTGCACCTTGATGCATACCGCGTGGTTCGCTTCGCTCGTGA 14087
456 eGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlap 473
14086 CAAAGAGCACAAAGATCCATCAGCGCGCTGAGT.....TAACCTCCAA 14043
473 roAlaGluLysArgSer.....GlnTyrProGluIleIlePheLeu 486
14042 AAGGTTCAGTCGCGAATGGTGCATATGAATCTTGAACGGTTATTPTTA 13993
487 GlyThrGlySerAlaIlePromethylsIleArgAsnValSerAlaThrLe 503
13992 GGCTGGCGGCATGGTGCATTACCCCATCGCATCTGACTTCGTGTT 13943
503 uValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyT 520
13942 GTTGGCTGCTGAGGAGAGTTC...TTTCTGTTCGACGCTGGGGAAGGTA 13896
520 hrPheGlyGlnLeuCysArg...HisTyrGlyAspGlnValAspArgVal 535
13895 CCCAAGTGTCTGGCGGCTCTGAGTTACGCTCGGAAGAAA..... 13856
536 LeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspHisH 552
13855 .....ATCAGTGCCTATTTTCATCAGTCACCCCATGCAGACACAT 13815
552 sThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSerL 569
13814 CACTGGGCTGCCGGGCTTTTTCATCTCTTCTCAAGTTGCTCGCAGC. 13766
569 euGlyLysProLeuHisProLeuValValAlaProAsnGlnLeuLys 585
13765 .....GAACGGCTGTACATCATCGTCTCCCAAGAACTGCA 13730
586 AlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValLeuHisHisI 602
13729 GAGTATGTGGAACAGCGCAGGATTTTGGACATGTACATTAATACGA 13680
602 eSerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerP 619
13679 GATTATTGTC.....AAAGAGGTGATAGAAC 13654
619 roAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGlu 635
13653 CACAAGTG.....GTATACCGTGGTAAAGATTTCAG 13622
636 GluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGlyCysAl 652
13621 GTGCGCTGTTTGTCTG...GATCATACCAAGCGGTGTATGGGTACAC 13575
652 alaLeu..... 653

```

13574 GCTAGAGACGAGGATCGTCCCGGATCATTTGACCCGCGCGGCTCAGG 13525
 654 ..ValHisThr..... 656
 : : : : :
 13524 ATTTCAGCTTCCCTGTGGGCGCTGTGGTCTCAGTTACAGTCAGGGTT 13475
 656 656
 : : : : :
 13474 GCAGTCAGTCGCGCGAGGGGTGACAGTATATCTCTGACAGGTAATGGG 13425
 657SerGlyTrpLysValValTyrSerGlyAspThrMetProC 670
 : : : : :
 13424 ACCGGCGCTCCGGCGCTAAGGTGAGCTCTGCTACTGATACAAATATT 13375
 : : : : :
 670 ySGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuLeuHisGlu 686
 : : : : :
 13374 TGCAGTCCATTGACGCGAGGTGCGGAATTCGGATTCTTTGTGTGAG 13325
 : : : : :
 687 AlaThrLeuGluAspGlyLeuGluGluAlaValGluLysThrHis 703
 : : : : :
 13324 GGAATGTTTGAGAAAGGATGGAAGGATGCGACGAGAGAAAGCATAT 13275
 : : : : :
 703 rThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGluPheI 720
 : : : : :
 13274 GAGTCCGTGCGAGCCACGATACGCCGCGATGCGCTGTCGACTTA 13225
 : : : : :
 720 leuMetLeuAsnHisPheSerGlnArgTyr 729
 : : : : :
 13224 TGGCACTCATCCTACTATAGCCCTCGATAT 13196

seq_name: /cgnl_8/gcgdata/geneseq/geneseqn/NA1999.DAT.X13218

seq_documentation_block:

XX ID X13218 standard; DNA: 5830 BP.

XX AC X13218;

XX DT 19-MAR-1999 (first entry)

XX DE Enterococcus faecalis genome contig SEQ ID NO:281.

XX KW Enterococcus faecalis; contig; detection; Enterococcal infection;

XX OS vaccine; attenuation; computer readable medium; ds.

XX OS Enterococcus faecalis.

XX PN W09850555-A2.

XX PD 12-NOV-1998.

XX PF 04-MAY-1998; 98WO-US08985.

XX PR 14-NOV-1997; 97US-0066009.

XX PR 06-MAY-1997; 97US-0044031.

XX PR 16-MAY-1997; 97US-0046655.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Barash SC, Dillon PJ, Kunsch CA;

XX DR WPI; 1999-045171/04.

XX PT New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.

XX PS Claim 1; Page 1306-1309; 2084pp; English.

XX CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC X12938 to X1319 represent these nucleotide sequences which are primary
 CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.

XX
 SQ Sequence 5830 BP; 1930 A; 966 C; 1262 G; 1666 T; 6 other;

alignment_scores:

Quality: 187.50 Length: 288

Ratio: 1.210 Gaps: 12

Percent Similarity: 53.819 Percent Identity: 27.431

alignment_block:

US-09-434-382-2 x X13218 ..

Align seg 1/1 to: X13218 from: 1 to: 5830

482 GluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAs 498
 : : : : :
 342 GAAATACAATTTTAGACACAGCGCTGGTGTACACAGCAAAACATCGCAA 391
 : : : : :
 498 nValSerAlaThrLeuValAsnIleSerProAspThrSerLeu.....L 513
 : : : : :
 392 TGTACACAGGAATTCGGTTGAAATTTAGTAGAACAGAAATGCCAGTTTGGC 441
 : : : : :
 513 euLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGly 529
 : : : : :
 442 TCTTCGATTGTGGGAGGAGCT.....CAATG..... 470
 : : : : :
 530 AspGlnValAspArgValLeuGlyThr.....LeuAl 540
 : : : : :
 471CAATTTTAAAGTAGTATTTCGCCCAAGAAAAATGA 508
 : : : : :
 540 aAlaValPheValSerHisLeuHisAlaAspHisHisThrGlyLeuProS 557
 : : : : :
 509 AAAGATTTTATTACACATTTTACATGGGATCATATTTTGGTTTACCTG 558
 : : : : :
 557 erIleLeuLeuGlnArg..... 562
 : : : : :
 559 GTTTGTTAGTAGTCGTTCTTTTCAGGCGGGACAGAACCTTTAGAAATC 608
 : : : : :
 563GluArgAlaLeuAl 567
 : : : : :
 609 TATGGACCACTTGGGATTGCTGACTTTGTCAAGACTTCTTTACGGGTCTC 658
 : : : : :
 567 a.SerLeuGlyLysProLeuHis....ProLeuLeu.ValValAlaProAs 582
 : : : : :
 659 GCAATCCGGTTCCTCTCCACTGAAATTTATGAATAACGAAAGAA 708
 : : : : :
 582 nGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGluValL 599
 : : : : :
 709 TGAGCTATTTTAAAGATATAACAATTCATCTGTCGCTGAATATCTTGG 758
 : : : : :
 599 euHisHisIleSerMetIleProAlaLysCysLeuGlnGlyAlaGlu 615
 : : : : :
 759 ACCATGCATCACGAGCTTTGGCTATCGA...ATTGAGGAAGCAGCGCAT 805
 : : : : :
 616 IleSerSerProAlaValGluArgLeu.....IleSerSerLe 628
 : : : : :
 806 GAAGGAGAAWTTMCAAGTAGAAGAACTTCAAGCTTTAGGATTCCTCTGG 855
 : : : : :
 628 uLeuArgThrCysAspLeuGluPheGlnThrCysLeuValArgHisC 645
 : : : : :
 856 GCCGTGTATGGCAACTCAACGCTGGTGAACG...ATTGTTTTCATG 902
 : : : : :
 645 ysLysHisAlaPheGlyCysAlaLeuVal.....HisThrSerGlyTrp 659

PR	28-JUN-1995;	95US-0000589.
XX	(IMCR) IMPERIAL CANCER RES TECHNOLOGY.	
PA	(UYUA) UNIV YALE.	
XX		
PA	Artavanis-Tsakonas S, Gray GE, Henrique D, Ish-Horowicz D;	
PI	Lewis J;	
PT		
XX		
DR	WFI; 1997-100159/09.	
DR	P-PSDB; W11720.	
XX		
PT	New vertebrate Delta protein, DNA and antibodies,- for treating and	
PT	preventing cancer, nervous system disorders and for tissue	
PT	regeneration	
XX		
XX	Disclosure; Fig 7A-7B; 135pp; English.	
XX		
CC	The M-delta-1 gene (#58899) codes for the mouse homologue (W11720)	
CC	of Drosophila Delta, a protein that binds to Notch protein. It was	
CC	obtd. by PCR amplification of cDNA derived from 8.5 to 9.5-day	
CC	mouse embryos using primers (see also T59457-58) based on Delta	
CC	and EGF-like repeat sequences. M-Delta-1 is primarily expressed	
CC	in the presomitic mesoderm, central nervous system, peripheral	
CC	nervous system, and kidney. Chick (#5897-98) and human (#58900,	
CC	T59454) Delta-1 sequences have also been isolated. Delta-1 genes	
CC	can be used in the prodn. of Delta polypeptides and (including	
CC	antisense sequences) utilised in the treatment of disorders of cell	
CC	fate or differentiation, such as cancer, and nervous system	
CC	disorders, or to promote tissue regeneration and repair.	
XX		
SQ	Sequence 2692 BP; 592 A; 748 C; 768 G; 581 T; 3 other;	

alignment_scores:		
Quality:	129.00	Length: 712
Ratio:	0.445	Gaps: 36
Percent Similarity:	40.730	Percent Identity: 20.927

alignment_block:

Align seq 1/1 to reverse of: T58899 from: 1 to: 2692

207 LeuSerProGlu.....ArgSerSerAspSerGluSerAsnGluAs 220

1749 CTTAGCCGGACGCAGACCACACAGCAGCAGCCAGCAGCAGGA 1700

220 nC|uBroHis|eouBro HisC|uVa|SerC|nArArC|uVa|A 235

```

z20 noiauoioiuieeafio...niiseiyvutseioinuayguyvata z33
      :::: ||| ||| ||| |||

```

1699 GGACAAGCACCAACCCCGGCACACACGGC..... 1672

235 rgAspSerLeuValValAlaIleHisLeuIysArg 251

$\vdots \vdots$ $||$
 $\vdots \vdots$ $||$

1671CACCCAGGGGAA 1660

252 GlyAsnPheLeuValLeuLys..... 258

[illegible]

1659 GGGCCCGCCCTGGCTCTCCATATGCCCTCTCACTGAGGTCACCAACCATGG 1610

259AlaLysGluMetGlyLeuProValGlyThrAla.... 269

[illegible]

TG09 GCGCIGGTGGTGGCTCAGGGAGGCAGAAATGGCAGTTGGGGCCGCCCATAG T380

270AlaIleAlaProIleIleAlaAlaVal 278

1559 CCTGGGGCGACACACATGCTAGCGCTGCGCGCGCTGCGCGCA
||:::|||||:::||||

[illegible]

279 LysAspGlyLysSerIleThrHisGluGlyArgGluIleLeuAlaGluG1 295

1492 CATGCTCACACCTGCTGACAGAGGGCGCTGCAGTCTCTT. 1455
312 roAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheGlnArg 328
1455GCCCCGTGTAGCCAGC 1441
329 TyrGlnGlyLysAlaAspAlaProValAlaLeuValValHisMetAlaPr 345
1440 TGGCGAGGTACAGGAGAA.GTCGTTCCACTGTCCTCC 1406
345 oAlaSerValLeuValAspSerArgTyrGlnGlnTrpMetClnuArgPheG 362
1405 GCGAGGTGCCCCCAATT.TGCACACAGGGGAG 1377
362 LyrAspThrGlnHisLeuValLeu. 370
1376 GAGGCACAGTCATCCACATTTGCTTCGCAGTACTCTCCGGAGAGCCAGC 1327
371AsnGluAsnCysAlaSerValHisAsnLeuArgSerHi 383
1326 CTGGCACCGGCACAGGTAAAGATTGCCGAGGTCCACACACTTGGCACCGT 1277
383 sLysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuL 400
1276 TAGAACAAAGGGAGAGCGCGCAGAGATCCAT.CTTCTTC 1239
400 euThrSerPheArgCysLysGluGlyProThrLeuSerValProMet 416
1238 TCACAGTTGAGCCAGAGAGCCCGAGGGCAATGGCAGGTGTAGCCCTCC 1189
417 ValGln. 418
1188 GTCAGGGTTATCTGAACATCGTCTCCATTGAAGCAAGGCCATCTGCAC 1139
419GlyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluT 433
1138 AGGTCATGGCGCTCAGCTCCACAGACCTTGCCTATAGAACGGGAGGCGAC 1089
433 rpGlnArgAspAlaIleThrCysAsnProGluGluPheIleValGlu 449
1088 GTGCAGAGAGAGCTGCTCAAGGTCCGTGCAGCTCCGCTCGTTCCTTGCA 1039
450 AlaLeuGln.LeuProAsnPheGlnSerValGlnGlu...Ty 463
1038 GGGCTAGGAGCACACTCATCTCTCCAGTCACAGTGTGGCACCTGTAT 989
463 rArgArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnT 480
988 ACCCAGGTGGCAGGAACATGTAGTCTCCCTGGCCGCT. 949
480 yrProGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIle 496
948GTTGGTCAGGTGGCTCCATTCCTGCACGGCTT 916
497 ArgAsnValSerAlaThrLeuValAsnIle.SerPr 508
915ATGGTGAGTACAGTAGTTCCAGTCTTGGTTCGACAAAGGCCCC 872
508 oAspThrSerLeu.LeuLeu. AspCysGlyGlu 518
871 CCGACCCCTTCCTGGCAGGTATACACTGCCAGGGTTGCTGGCAGGTCCCATGG 822
518 uGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgV 535
821 AGACAACCTGGGTATFCGGATGCATCTACGCGAGTAGCGGCCCTGCCAGCC 772
535 aIleGlyThrLeuAlaValPheValSerHisLeuHisAlaAspHis 551
771' AACTCTGCACCTGCACCTCCCTGGCTTTGTCACAGTATCCATGTGGTCAT 722
552 HisThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlase 568
721 CACACC...CTGGCAGCAGATTTGGGTGCTGAGTCAGTACTGGCTCTCCAG 675

[illegible]

```
seq_name: /cgnl_8/gcgdata/geneseq/geneseq/NA1997.DAT.x30827
seq_documentation_block:
ID x30827 standard: DNA: 551 BP.
```


alignment_scores:

Quality: 123.00 Length: 939
 Ratio: 0.330 Gaps: 48
 Percent Similarity: 39.723 Percent Identity: 19.276

alignment_block:

US-09-434-382-2 x V34766 ..

Align seg 1/1 to: V34766 from: 1 to: 9759

```

2 TrpAlaLeuCys.....SerLeuLeuArgSe 10
||| |||||
3943 TGCACCTGTGCGCGCACCGACGAGGGGTGGCCCGAGGTACTACGACG 3992

10 rAlaAlaGly.ArgThrMetSerGlnGlyArgThrIleSerGlnAlaPro 26
||| |||
3993 ACCTCGAGGTGCGCGCTCGGGATGACGCCATGGCCCGGGCGGCCTC 4042

27 AlaArgGluArgProArgLysAspProLeuArgHisLeuArgThrAr 43
|||
4043 GCATCAGTCCACGCGCTCGCAAA..... 4066

43 gGluLysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrL 60
|||||
4067 .....GGCCCTTACAATATCAGGG 4085

60 euGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrVal 76
::
4086 TATGGAACTGCGCGGAGGC.....CCTGGCAAGACCACCGCGATC 4126

77 PheSerGluPheAsnArgTyr...LeuPheAsnCysGlyGluGlyValG 92
:::
4127 CTGCGTGCCTTACGCGGAGACCTTTACGTCTGC.....CCACCAA 4170

92 nArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIleP 109
:
4171 TGGCGTCTCTCAGCAGATCCAGCCCAAACTCCGCGCGCGGAT..... 4213

109 heLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
4213 ..... 4213

126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProP 142
::
4214 ATCGAGATCAAGAAGCGCGCCACTACGAGCGCGCTGACGAACCGCT 4263

142 oGlnLeuLys.....TyrLeuGluAlaIleLysIlePheSerGly. 156
::
4264 CGCGCGCTACCGCGCATCTACATCATGATGAGGGTTTCACTCTCGGGGCG 4313

157 .....ProLeuLysGlyIleGluLeu 163
4314 AGTACTGCGGCTTCGTTGCCAGCAACACCGCGGAGGTGATCTGCGTC 4363

164 AlaValArgProHisSerAlaProGluTyrGluAspGluThrMetThrVa 180
::
4364 GGTGATCGGGACCAAGTGGCGCCACCACTACGCCAATAACTGCCGAC... 4411

180 lTyrGlnIleProIleHisSerGluGlnArgArgGlyLysHisGlnProt 197
::
4412 ....CCGTCCTCCGACCGTGGCTACCGAGCGCTCGCGCAC...ACTT 4454

197 rpGlnSerProGluArgProLeuSerArgLeuSerProGluArgSerSer 213
|||
4455 GGCCTTCCCGACTGCTGGCGGCGCGCTGCGCGCGGGGCTCGATTAT 4504

214 AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSerG 230
|||
4505 GACATCGAG.....GGCGA 4518

230 nArgArgGlyVal.....ArgAspSerSerL 239

```

```

4519 GCGCACCGGCACCTTCGCTGCAACCTTTGGGACGGCCCGAGTGCAC 4568
||| |||
239 euValValAlaPhe.....IleCysLysLeuHisLeuLysArg 251
||
4569 TTCACCTCGCTTCTCGCGGAAACCGTGGCGCGCTTAC....GAGGCT 4615

252 GlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyTh 268
|||
4616 GGCATACGCGCATACACCGTGGCGGAGGCCAGGTATGAGCGTCGGCAC 4665

268 rAlaAlaIle..... 271
|||
4666 CGCTGCATCCATGTAGGACAGACGACCGACGTTGCCCTGGCGCTGA 4715

272 .....AlaProIle 274
|||
4716 CACGCGACCTCGCATCGTCAGCCTGACCGCGGCTCCGACGCACTAC 4765

275 IleAlaAlaValLysAspGly..... 281
::
4766 CTCACGAGCTCGAGGACGCTCACTGCGCGCTGCGGGGCTCAGCGGTT 4815

282 .....LysSerIleThrHisGluGlyA 289
|||
4816 CCTCGACGCGCGGGCACCTGCGGAGCTCAAGGAGTTTCCGCTGGCATTTG 4865

289 rgGluIleLeuAlaGluGluLeuCysThrProProAspProGlyAlaAla 305
::
4866 ACCGCGTGTGCGCGTGCAGCAGCAGCACCACCGTTCGCGCGCGCGCAC 4915

306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysG 322
::
4916 GGCATC.....CCGAGGCCAACAGCTGCGCGCTTCTGCC 4953

322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAla 339
::
4954 CCGCACTCTGGAGGAGCTCTCTTCGCGCTGCGCGCACCC... 4996

339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
|||
4997 .....CATTACGCGGACCTCAACCGCTGACTGAGGCGGACGAGAA 5038

356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnG 372
|||
5039 .....GTGCGGTATATGCGCATCTCGCGTCAACCTG...CTCAACAA 5076

372 uAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGln 389
|||
5077 GAATCACACCGAGATGCCGGA.....ACGGAACGCGTTC 5111

389 euAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgCys 405
|||
5112 TCAGTGGCGTTTCGCCG.....TCGCGCTACCGCGCG 5143

406 LysLysGluGlyProThrLeuSerValProMetValGlnGlyGluCysLe 422
|||
5144 GCGAGGATGGTTCGACCTCCGCACTGCTGGCGCGCGCAGCACCAGCG 5193

422 uLeuLysTyrGlnLeuArgProArgArg.....G 432
|||
5194 CCCTTTTCGCGCAGATCCACCGCGCGCTCACTGCTGGGGTGGCGCAGG 5243

432 luTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleVal 448
|||
5244 AGTGGCGCATG.....ACGTACTTGGCG 5266

449 GluAlaLeuGlnLeuProAsnPheGlnIleSerValGlnGluTyrArgAr 465
|||
5267 GAACGATCGACCTCACTGACGTCTACACGACAGATGGCGTGGCGCGCG 5316

465 gSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProG 482
|||

```

```

724 .HisPheSerGlnArgTyrAlaLysValProLeuPheSerProAsnPhe 739
||||| : ::::::::::::::||| ::::::::::|||
6018 TCACATCCCGGTGAAGCATGTGAGCACCCCT .....ACCCCAGCTTC 6061
740 SerGluLysValcIyValalalpheAspHisMetLysvalCysPheGlyAs 756
||||| : ::::::::::||| ||| |||
6062 TCGGGGCACCTCGGCACCGCGCGCGC.....CTCTTCCATGA 6099
756 pPheProThrMetProLysLeuIleProProLeuLysAlaLeuPheAlaG 773
||||| : :::: ||| ||| |||
6100 T.....GTCACTCACCGACGCGATCAAGTGCTTTGCCGCCGCTTCGACC 6143
773 lyAspfileGluMetGluGluArgArgGluLysArgGluLeuArgGln 789
||||| : ||| : :::: ::::::::::
6144 CAGACGTGCTTTGAAGAAGACAGCAGGTGGCCCTCTCTCGACCGCTCCGGGG 6193
790 ValArgAlaAlaLeu 794
||| ||| ||| |||
6194 GTCTACGCGGCTCTG 6208

seq_name: /cgml_8/gcddata/geneseq/geneseqn/NA1995.DAT:Q90529
seq_documentation_block:
ID Q90529 standard; cDNA; 3501 BP.
XX XX
AC Q90529;
XX XX
DT 02-NOV-1995 (first entry)
XX XX
DE Rat SIII 110 kDa subunit cDNA.
XX XX
KW RNA-polymerase transcription factor SIII; ss.
XX OS Rattus sp.
XX XX
FH Key Location/Qualifiers
CD S 82..2403
FT FT /*tag= a
FT misc_difference 2990
FT /*tag= b
FT /*note= "base 2990 can be a, c, t or g"
XX XX
EP E65498-A.
XX XX
PD 31-MAY-1995.
XX XX
PF 29-NOV-1994; 94EP-0250286.
XX XX
PR 30-NOV-1993; 93US-0160087.
XX PA (OKLA-) OKLAHOMA MED RES FOUND.
XX PI Bradsher JN, Conaway JW, Conaway KC;
XX WPI; 1995-195587/26.
DR P-PSDB; R75358.
XX XX
PT Novel RNA polymerase transcription factor SIII - useful for
PT the rate of transcription by RNA polymerase II.
XX Claim 29; Page 21-24; 43pp; English.
XX CC
XX Trypsin peptides (given in R75096-97) of purified rat SIII
CC subunit were used to design probes (Q90538-39) used to screen
CC a rat brain lambda-CEM2 cDNA library. Hybridizing clones were
CC further screened by PCR using primers (Q90539-40). cDNA
CC (Q90529) encoding the p110 protein (R75358) was obtained.
XX SQ Sequence 3501 BP; 984 A; 860 C; 896 G; 761 T; 0 other;
```

alignment_scores:

Align seg 1/1 to: Q90529 from: 1 to: 3501

237 rSerLeuValValAlaPheIleCysLysLeuHisLeuLysArgGlyAsnP 254
::: : ::|||::: |
681 GGAACCCATCGTCTCA.....CACCCGAAGCTGGGAAG 715

254 heLeuVal.....LeuLysAlaLysGluMetGlyLeu 264
||| : ::|||::: |||
716 TCCACAGTAATACCTTTTCAGGACAGACTAGGGGTAGTACCCTGGGTAG 765

265 ProValGlyThrAlaAlaIleAla..... 272
||| : ::|||::: |
766 CACCANGGANAGGGCTGTAGCCAANAACGCCACAANACTTCCCCA 815

273ProIleLeuAlaValLysAspGlySerIle. 284
|||::: ||| : ::|||::: |
816 CAAGGAGAAACGCCCTGTGGATGCCAGGGGATGAGAAGACTCTCTCA 865

285ThrHisGluGlyArgGlu 290
::: |||::|||::: |
866 TGGCAGAGAGAAAGTCAACAAGACCTTTCCAAGAGGAGAGCCGAAG 915

291 IleLeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheVa 307
::: |||::: |
916 CTACTCTCAGAGACAGTGCCCAAGGAGAAGCTGCCCTCCAGT.....GT 959

307 lValValGluCysProAspGluSerPheIleGlnProIleCysGluAsnA 324
||| : ::|||::: |
960 TGTCAAAGAAGAAGAGACAGA..... 981

324 laThrPheGlnArgTyrcInGlyLysAlaAspAlaProValAlaLeuVal 340
::: |||::: |
982GAAGGC AAC.....AGCCTCAAG 999

341 ValHisMetalProAla...SerValLeuValAspSerArgTyrcInGl 356
::: |||::: |
1000 AAGAGTGTGCACCTGCCCTTAGATTGTTCTTCACACAACCACTTTAAAAA 1049

356 nTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluA 373
: : |||::: |
1050 GCCCAAACACAAAG.....GACTCCGAAAATAACAAGTCTGACAAAA 1090

373 snCysAlaSerVal..... 377
|| : |||||
1091 ACAAGCAGAGTGTAGATAGCGTGGACTCAGGACGAGGACAGGAGACCCA 1140

378HisAsnLeuArgSerHis..... 383
::: |||::|||::: |
1141 TTACCCAGACCAAGGATAAAGTTTCCCAACNACCTGAAGCTCAGGAGGG 1190

384 ..LysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuL 400
::: |||::: |
1191 GAAAGTAAGAATAACTAATCGGGATGAAAAGTCAACAGGCTCACTCCCTAAAG 1240

400 euThrSerPheArgCys.....LysLysGluGlyProThrLeuSerVal 414
: : : : :
1241 TAGAAGAGATGGACATGGATGATGAGTTGAGCAGCCCACCATGCTCCTTT 1290

415 ProMetValGlnGlyGluCysLeuLeuLysTyrcInLeuArgProArgAr 431
||| : |||::: |
1291GAGTCATACCTCAGCTATGAC...CAGCCCCGCCAA 1322

431 gGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleV 448
: : : : :
1323 G.....AAAAGAAGAAGGTTGTGAAAACCTCCGGTACAGCAGCTTGGAG 1366

448 alGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrcArg 464
: : : : :
1367 AAAAGGACTTTAAAAAGAAGAC...TCTAAAGACACTAGTAAAAACTTG 1413

465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrcPr 481
: : : : :
1414 ANCTCGCTCAGAAATTATCCCAAGCGGNATGAAACCAAGTCAGACAG.. 1461

481 ocLuIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498

Search time (sec): 77.370000

100

476 LysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySerAlaIle 492

```
2154 AAAGAGGGCCATATATGGAAGTTACATTTTGTGAACGAGTGCGAGTTT 2203
492 ePrometLysileArgAsnValSerAlaThrLeuValAsnIleSerPro. 508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2204 GCCTACAAAGAGAGAAATACACAAAGCAATCGCCTTAATTTAGAACCAT 2253
509 .....AspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGln 523
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2254 ATTCCAATTCCATATGCTTTTCGACGTTGGTGAAGGTACACAGCACCAA 2303
524 LeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAl 540
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2304 ATTTTACATCAT.....GCAATTAAATTAGGAAAGTGAC 2338
540 aalaValPheValSerHisLeuHisAlaAspHisThrGlyLeuProS 557
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2339 ACATATATTTATTACTCATATGCGATGCGATCATATTTTGGTTGCCAG 2388
557 erIleLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeu 573
    : : : : : : : : : : : : : : : : : : : : : : : : : :
2389 GATTACTTCTAGTCGT.....TCTTTTCAGGGCGGTGAACAG 2426
574 HisProLeuValValAlaProAsnGlnLeuLysAlaTTPLeuGlnGl 590
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2427 AAGCGCTTACATGTTGGACCAAGGAATTAAAGCATATGTGGAAAT 2476
590 nTyrHisAsn 593
    : |||
2477 GTCATGAAT 2486
```

seq_name: /cgnl_7/ptodata/1/ina/5A_COMB.seq:us-08-524-757-5

```
seq_documentation_block:
: Sequence 5, Application US/08524757
: Patent No. 5792634
: GENERAL INFORMATION:
: APPLICANT: Conaway, Ronald C.
: APPLICANT: Conaway, Joan W.
: APPLICANT: Bradsher, John N.
: TITLE OF INVENTION: RNA Polymerase Transcription Factor
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
: STREET: 1201 Elm Street, Suite 4500
: CITY: Dallas
: STATE: TX
: COUNTRY: US
: ZIP: 75270-2197
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/524,757
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/13621
: FILING DATE: 29-NOV-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/160087
: FILING DATE: 30-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Harre, John A.
: REGISTRATION NUMBER: 37,345
: REFERENCE/DOCKET NUMBER: B35006CIPCIP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (214) 939-4500
: TELEFAX: (214) 939-4600
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3501 base pairs
```

```
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: both
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 82..2403
: FEATURE:
: NAME/KEY: misc.difference
: LOCATION: replace(2990, "")
: OTHER INFORMATION: /note= "This base can be either a,
: OTHER INFORMATION: c, t, g."
US-08-524-757-5
```

```
alignment_scores:
Quality: 120.50 Length: 763
Ratio: 0.378 Gaps: 38
Percent Similarity: 41.809 Percent Identity: 19.528
alignment_block:
US-09-434-382-2 x US-08-524-757-5
```

```
Align seg 1/1 to: US-08-524-757-5 from: 1 to: 3501
13 GlyArgThrMetSerGlnGlyThrIleSerGlnAlaProAlaArgArg 29
||| : : : : : ||| ||| : : : : : |||
14 GCGGAGTTGCGGCGCGGAGGACGCTACGCGGCCAGCTTCGCGCGAGGAG 63
29 gGluArg.....ProArgLysAspProLeuArgHisLeuArg.ThrArg 43
||| ||| ||| ||| ||| ||| ||| ||| |||
64 GCCGCGCAGTGACAGCGCTGCGCGGAGTGGCGGCTCCCAAGTTGTGA 113
44 GluLysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrLe 60
||| ||| ||| ||| ||| ||| ||| ||| |||
114 GAAGCTGACGGCGCGCTGGCTGGCAACCGGACCCCAA..... 152
60 uGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrValP 77
153 .....GAGCTATTGAAATAT 168
77 heSerGluPheAsnArgTyrLeuPhe.....AsnCysGlyGlu 89
||| ||| ||| ||| ||| ||| ||| ||| |||
169 TTGAAGAACTTTCGCTTACCTATTACGTAGACATTTCTGTGGAGAC 218
90 .GlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuA 106
||||| : : : : :
219 TGGGTGGGAAACAGTG..... 237
106 spAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSer 122
||| ||| ||| ||| ||| ||| ||| ||| |||
238 ..AACAGCTTT.....CGGAACAT...GAGCAAGTGGGAACCTTGCC 276
123 GlyMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValLeuSe 139
: : : : : : : : : : :
277 AGAGACCTGTTGCGCAGTGGGAAGAGCTGGTTCCA..... 312
139 rGlyProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSerG 156
: : : : : ||| |||
313 .....GTAGAACGAAACAAATGAGGCT.....GAGGATC 340
156 lyProLeuLysGlyIleGluLeuAlaValArgProHisSerAlaProGlu 172
: : : : : ||| ||| ||| ||| ||| ||| ||| |||
341 AGGATTTTGAGAGAGCAATTCGCCCAAGCGTCCCGAGATGTTCCCCAG 390
173 TyrGluAspGluThrMetThrValTyrGln..... 182
||| ||| ||| ||| ||| ||| ||| ||| |||
391 CAGGAGGAGGAAGCGAGGGGAACTACCAAGGAAAGCTGGCAAGCCTCAGG 440
183 ....IleProIleHisSerGluGlnArgArgGlyLysHisGlnProTpg 198
||| ||| ||| ||| ||| ||| ||| ||| |||
441 CAGCCAGCCCTACAGCCCTGAGCACAGACAGAAAAAGCACAGAAAACTTC 490
```



```

198 InSerProGluArgPro..... 203
491 CTGAGCTTGAAGGCGCTCACAAGTGGGTCTACGTCACGAGGAGAGAT 540
204 .....LeuSerArgLeuSerProGluArgSerSerAspSerG1 216
541 GAAAGGAAGAGGTGTCACAAAGGTGTCCACCACCATATTCTTCAGACCCCGA 590
216 uSerAsnGlu.....AsnGluProHisL 224
591 GTCGCTGCTGACTATGTCATGTTCAATCTCTCCACCTTCAAGTCCCCATC 640
224 euProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuVal 240
641 AAATGATATACAGACCTCTCTAGTCCCGAGATGGACACGAGAACCCATC 690
241 ValAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuVal... 256
691 GTCTCA.....CACCGAAGCCTGGGAAGTCCACAGTAA 725
257 .....LeuLysAlaLysGluMetGlyLeuProValGlyT 268
726 TACCTTTTCAGACACACTAGGGGTAGTCACCTGGGTGACCAAGGA 775
268 hrAlaAlaIleala..... 272
776 AAGGGCTGTTAGCCAAACCAAGCCACAAATCTCCCAAGGAGAA 825
273 ...ProIleIleAlaAlaValLysAspGlyLysSerIle..... 284
826 CCGCTGTGGATGCCAGGGGGATGAGAGACTCTGTCAATGGGAGAGA 875
285 .....ThrHisGluGlyArgGluIleLeuAlaG 294
876 GAAGTCACAAAGCCTCTTCCAAAGAGAGAGCGGAGGCTACTCTCAG 925
294 luGluLeuCysThrProProAspProGlyAlaAlaPheValValGlu 310
926 AGGACAGTGCACAGGAGAGAGCTGCCCTCCAGT.....GTTCTCAAGAAA 969
311 CysProAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheG1 327
970 GAGAGGACAGA..... 981
327 nArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValHisMetA 344
982 .....GAAGGCAAC.....AGCCTCAAGAAGAAGTTGT 1009
344 laProAla...SerValLeuValAspSerArgTyrGlnGlnTrpMetGlu 359
1010 CACCTGCCTTAGATGTTGCTTCAGACACACCACTTTAAAAAGCCCAACAC 1059
360 ArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCysAlaSe 376
1060 ARG.....GACTCCGAGAAATCAAGTCTGACAAAAACACGACAG 1100
376 rVal..... 377
1101 TGTACATAGCGTGACTCAGGACGAGGACAGGAGACCACTTACCAGAG 1150
378 .....HisAsnLeuArgSerHis.....LysIleGln 386
1151 CCAAGGATAAAGTTCACCAACACCTGAAGGCTCAGAGGGGGAAGTAAGA 1200
387 ThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPh 403
1201 ACTAACTGGATCGAAGTACACGGCTCACCTCCCTAAAGTAGAAGAGAT 1250
403 eArgCys.....LysLysGluGlyProThrLeuSerValProMetValG 418
1251 GGACATGGATGATGAGTTGACGACCCCAACCATGTCCTTT..... 1290
418 lnglyGluCysLeuLeuLysTyrGlnLeuArgProArgArgGluTrpGln 434

```

```

1291 ....GAGTCATACCTCAGCTATGAC...CAGCCCGCAG.....AAA 1326
435 ArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAlaLe 451
1327 AGAAGAAGGTTGTGNAAACTTCGGGTACAGCACTTGGAGAAAAAGGACT 1376
451 uGlnLeuProAsnPheGlnSerValGlnGluTyrArgArgSerAlaG 468
1377 TAAAAAGAAGGAC...TCTAAAAGCACTAGTAAAAAATTGAACCTCGCTC 1423
468 InAspGlyProAlaGluArgSerGlnTyrProGluIleIle 484
1424 AGAAATTACCAAGCGGAATGAACAACAGTCACAGAAG..... 1461
485 PheLeuGlyThrGlySerAlaIlePrometLysIleArgAsnValSerAl 501
1462 ...TTGCACCGAGCTGGAGCCGACCCAGAGGCTTAGAAGGTCCCTAC 1508
501 aThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyG 518
1509 TGATGTGCTGCCAGCATTCACAGACATC..... 1536
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArg 534
1536 ..... 1536
535 ValLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspHi 551
1537 .....CCCTTGCCCGCCATA.....CAAACCAA 1559
551 shisThrGlyLeuProSerIleLeuLeu..... 560
1560 CTATCGTCCCTTCCCTCCCTCGAGTTGATCTCCTCTTTCACCCAAAGC 1609
561 .....GlnArgGluArgAlaLeuAlaSerLeuGly 570
1610 GAAAAGCTTCTCTTCAACCCAGGAGAAAGAAAGCTGGGTTCACAGGA 1659
571 LysProLeuHisProLeuLeuValValAlaProAsnGlnLeuLysAlaTr 587
1660 CCGAGNATGAATCTAAGATGCGAGGTGATTTCAGGTTCCAGTGCCTA 1709
587 pLeuGlnGln.....TyrHisAsnGlnCysGlnGluValLeuHisH 601
1710 TCTCCCCAAATGATGACCTTGACACGAGTATATCCGGGTGCTTAAGA 1759
601 is.....IleSerMetIleProAlaLysCysLeu 610
1760 ATAATATTGACTCCATCTTTGAAGTGGGAGGATCCCTTATTCTGTCTT 1809
611 GlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerSe 627
1810 GAA.....CCTGCTTGGAGAGGTGCACACCCGA 1838
627 rLeuLeu.....ArgThrCysAsp.....LeuGluGluPheG 638
1839 TCAGCTCTACCGAATAGGAATGCAATCATGTATTAAATTGAGGAAACAG 1888
638 InThrCysLeuValArgHisCysLysHisAlaPhe 649
1889 ATCAGTTGTGAAAGTTCACTGTCACCGGAGCTTT 1923

```

seq_name: /cgnl_7/ptodata/1/ina/5A_COMB.seq:US-08-459-041A-1

seq_documentation_block:

; Sequence 1, Application US/08459041A

; Patent No. 5663065

; GENERAL INFORMATION:

; APPLICANT: Frey, Teryl K.

; APPLICANT: Dominguez, Geraldina

; APPLICANT: Wang, Chin-Yen

; TITLE OF INVENTION: DNA Encoding Infectious Rubella Virus

```

:
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Jones & Askew
: STREET: 191 Peachtree Street, 37th Floor
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/459,041A
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Greene, Jamie L.
: REGISTRATION NUMBER: 32,467
: REFERENCE/DOCKET NUMBER: 07362-0102
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 818-3700
: TELEFAX: (404) 818-3799
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9759 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-459-041A-1
:
: alignment_scores:
: Quality: 120.00 Length: 939
: Ratio: 0.323 Gaps: 48
: Percent Similarity: 39.617 Percent Identity: 19.276
:
: alignment_block:
: US-09-434-382-2 x US-08-459-041A-1 ..
: Align seg 1/1 to: US-08-459-041A-1 from: 1 to: 9759
:
: 2 TrpAlaLeuCys.....SerLeuLeuArgSe 10
: ||| |||||
: 3943 TGGACCTGTGCCCCGACCGACGAGGGCTGGCCCGGCTACTACGACG 3992
:
: 10 rAlaAlaGly.ArgThrMetSerGlnGlyArgThrIleSerGlnAlaPro 26
: ||| ||| :|||:
: 3993 ACCTCGAGGTGCGCGCTCGGGGATGAGCGCATGGCCCGGCGCCCTC 4042
:
: 27 AlaArgArgGluArgProArgLysAspProLeuArgHisLeuArgThrAr 43
: ||| :|||:|||||
: 4043 GCATCACTCAACCGCCCTCGCAA..... 4066
:
: 43 gGluLysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrL 60
: ||||| :|||:
: 4067 .....GGCCCTTACATATACAGG 4085
:
: 60 euGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaLeuTyrVal 76
: :|||:|||||
: 4086 TATGAACATGGCCGACGC.....GCTGGCAAGACTTACCCGCATC 4126
:
: 77 PheSerGluPheAsnArgTyr...LeuPheAsnCysGlyGluGlyValG 92
: :|||:|||||
: 4127 CTCGCTGCCCTTCAGCGCGAAGACCTTTACGTCGC.....CCACCAA 4170
:
: 92 nArgLeuMetGlnGluHisLeuLysValAlaAlaArgLeuAspAsnIleP 109
: |||:|||||
: 4171 TGGCCTCTCGACGAGATCCAGGCCAACCTCCGCGCGCGCGAT..... 4213

```

```

109 heLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
4213 ..... 4213
126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProPr 142
:||||:|||||
4214 ATCGACATCAAGACGCGCCACCTACGAGCGCGGCTGACGAAACCGCT 4263
142 oGlnLeuGluLys.....TyrLeuGluAlaIleLysIlePheSerGly. 156
:||||:|||||
4264 CGCCGCTACCGCCCATCTACATCGCATGAGCGGTTCACTCTCGCGCGCG 4313
157 .....ProLeuLysGlyIleGluLeu 163
4314 AGTACTGCGCGTTGTTGCCAGCCAAACACCGGAGGTGATCTCGCTC 4363
164 AlaValArgProHisSerAlaProGluTyrGluAspGluThrMetThrVa 180
:||||:|||||
4364 GGTGATCGGACAGTAGTGGCGCCACACTACGCCAATAACTGCCGACCC.. 4411
180 lTyrGlnIleProIleHisSerGluGlnArgGlyLysHisGlnProT 197
:||||:|||||
4412 ....CCGTCCTCACCCTGGCTACCGAGCGCTCGGCCAC...ACTT 4454
197 rpGlnSerProGluArgProLeuSerArgLeuSerProGluArgSerSer 213
:||||:|||||
4455 GCGCTTCCCGACTGCTGGCGCGCCGCTCGCGCGGCTCGATTAT 4504
214 AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSerGl 230
:||||:|||||
4505 GACATCGAG.....GGCGA 4518
230 nArgArgGlyVal.....ArgAspSerSerL 239
:||||:|||||
4519 GCGACCGGCACCTTCGCTCAACCTTTGGAGCGCGCCGAGTCTGACC 4568
239 euValValAlaPhe.....IleCysLysLeuHisLeuLysArg 251
|| :|||||
4569 TTCACCTCGCTTCGCGCGGAAACCGGTGCGCCCTTAC...GAGGCT 4615
252 GlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyTh 268
:||||:|||||
4616 GGCATACGCGCATACACCGTCCGCGAGGCCAGGTATGAGCGTCGGCAC 4665
268 rAlaAlaIle..... 271
:||||:|||||
4666 CGCTGATCATCTAGCAGAGACGCGACGCTTGCCTTGCCTGCGCTGA 4715
272 .....AlaProIle 274
:||||:|||||
4716 CACGCGACCTCGCCATCGTCAGCCTGACCGGGCTCCGAGCGACTCTAC 4765
275 lIleAlaValLysAspGly..... 281
:||||:|||||
4766 CTCACGAGCTCGAGGACGCTACTCGCGCTCGGGGCTCAGCGCGTT 4815
282 .....LysSerIleThrHisGluGlyA 289
:||||:|||||
4816 CCTCGACCGCGGCGACTGGCGGAGCTCAAGGAGGTCCCGCTGGCATTTG 4865
289 rgGluIleLeuAlaGluLeuCysThrProProAspProGlyAlaAla 305
:||||:|||||
4866 ACCCGGTGTGCGCGTCGAGCAGCAGCACCACCGCTTGGCGCGCGCGCAC 4915
306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysGl 322
:||||:|||||
4916 GGCATC.....CCGAGGCCCAAGACGTCGCGCGCTTCTGCC 4953
322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
:||||:|||||
4954 CCGCATCTCGAGGAGCTGCTCTTCGCGCGGTGCGCGGCCACCCC..... 4996

```



```

; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/999,733
; FILING DATE: 02-SEP-1997
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/459,041
; FILING DATE: 02-JUN-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/093,453
; FILING DATE: 19-JUL-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/722,334
; FILING DATE: 28-JUN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Greene, Jamie L.
; REGISTRATION NUMBER: 32,467
; REFERENCE/DOCKET NUMBER: 07362-0103
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 818-3700
; TELEFAX: (404) 818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9759 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-999-733-1

alignment_scores:
  Quality: 120.00      Length: 939
  Ratio: 0.323         Gaps: 48
  Percent Similarity: 39.617      Percent Identity: 19.276

alignment_block:
  US-09-434-382-2 x US-08-999-733-1
  Align seg 1/1 to: US-08-999-733-1 from: 1 to: 9759

2 TrpAlaLeuCys.....SerLeuLeuArgSe 10
||| |||||
3943 TGGCACCTGTGCGCGACGACGAGGGCTGGCCCGCGTACTACGAGC 3992
10 rAlaAlaGly ArgThrMetSerGlnGlyArgThrIleSerGlnAlaPro 26
||| ||| .....
3993 ACCTGAGGTGGCGGCTCGGGATGACGCCATGGCCGGCGGCGCCCTC 4042
||| .....
4043 GCATCAGTCCAAACGCCCTCGCAAA..... 4066
43 gLulysArgGlyProSerGlyCysSerGlyGlyProAsnThrValTyrL 60
.....GGCCCTTACAATATCAGGG 4085
60 euGlnValValAlaAlaGlySerArgAspSerGlyAlaAlaTyrVal 76
.....
4086 TATGAACATGGCGCAGGC.....GCTGGCAGACTACCGCATC 4126
77 PheSerGluPheAsnArgTyr...LeuPheAsnCysGlyGluGlyValG1 92
.....
4127 CTCGCTGCCTTCACGCGCGGAAGACCTTTACGCTGC.....CCACCAA 4170
92 nArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspAsnIleP 109

```

```

; |||:|||||: |||: |||: |||
4171 TCGCTCTGTGACGAGATCCAGGGCAAACTCGCGCGCGCGAT..... 4213
109 heLeuThrArgMetHisTrpSerAsnValGlyGlyLeuSerGlyMetIle 125
4213 ..... 4213
126 LeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerGlyProPr 142
; |||:|||||: |||: |||: |||
4214 ATCGACATCAAGACGCGCGCACCTACGAGCGCGCGCTGACGAAACCGCT 4263
142 oGlnLeuGluLys.....TyrLeuGluAlaIleLysIlePheSerGly. 156
; |||:|||||: |||: |||: |||
4264 CGCGCCTTACCGCGCATCTATCATGAGGCGCTTCTCTCGCGCGG 4313
157 .....ProLeuLysGlyIleGluLeu 163
4314 AGTACTGCGCGCTTCTTCCAGCCAAACCCAGCGGAGGTGATCTGCGTC 4363
164 AlaValArgProHisSerAlaProGluTyrGluAspGluThrMetThrVa 180
; |||: |||: |||: |||: |||
4364 GGTATCGGACGAGTGGCGGCCACACTACGCCAATAACTGCCGACCC... 4411
180 lTyrGlnIleProLleHisSerGluGlnArgGlyLysHisGlnProT 197
; |||: |||: |||: |||: |||
4412 ....CCGCTCCTGACCGCTGGCCTACGAGCGCTCGCGCCAC...ACTT 4454
197 rpGlnSerProGluArgProLeuSerArgLeuSerProGluArgSerSer 213
; |||: |||: |||: |||: |||
4455 GCGCTTCCCGACTGTGGCGGCCCGCTCGCGCGGGGCTCGATTAT 4504
214 AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSerG1 230
||| ||| .....GGCGA 4518
4505 GACATCGAG.....ArgAspSerSerL 239
; |||: |||: |||: |||: |||
4519 CGCACCGGACCTTTCGCTGCAACCTTTGGGACGCGCGCGAGTCTGACC 4568
239 euValValAlaAlaPhe.....IleCysLysLeuHisLeuLysArg 251
|| .....
4569 TTCACCTCGCCTTTCGCGCGCAACCGCTGCGCGCGCTTCAC...GAGGCT 4615
252 GlyAsnPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyTh 268
||| .....
4616 GCATACGCGCATACACCGTGGCGGAGGCCAGGCTATGAGCGTGGCAC 4665
268 rAlaAlaIle..... 271
|||: |||: |||: |||: |||
4666 CGCCTGCATCCATGTAGCAGAGACGCGACGCTTCCCTGCGCGTGA 4715
272 .....AlaProIle 274
4716 CACGGACCTCGCATCTGTCAGCTGACCGCGCGCGCTCGCGACTCTTAC 4765
275 IleAlaAlaValLysAspGly..... 281
; |||: |||: |||: |||: |||
4766 CTCCACGAGCTCGAGGACGCTCTCTGCGCGCTGCGGGGCTCAGCGCTT 4815
282 .....LysSerIleThrHisGluGlyA 289
|||: |||: |||: |||: |||
4816 CCTCAGCGCGGGGACTGCGCGAGCTCAAGGAGGTTCCTCCGCTGGCATTG 4865
289 rgGluIleLeuAlaGluGluLeuCysThrProAspProGlyAlaAla 305
; |||: |||: |||: |||: |||
4866 ACCGGTGTGCGCGTTCGAGCAGGACGCGCGCTTCCCGCGCGCGAC 4915
306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysG1 322
; |||: |||: |||: |||: |||
4916 GGCATC.....CCGAGGCGCGAGACGCTGCGCGCTTCTGCGC 4953
322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
; |||: |||: |||: |||: |||

```



```

; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036,987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 80161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-036-987A-1

```

```

alignment_scores:
  Quality: 119.00      Length: 749
  Ratio: 0.378        Gaps: 50
  Percent Similarity: 42.056  Percent Identity: 21.762

alignment_block:
  US-09-434-382-2 x US-09-036-987A-1/rev ..

Align seg 1/1 to reverse of: US-09-036-987A-1 from: 1 to: 80161

      8 LeuArgSer.....AlaAlaGlyArgThrMetSerGlnGlyAr 20
      |||||
      57302 CTCGAAGCGGACAGGATCACGCACTTGGCGCTCCAGTAGTCGGCGCG 57253

      20 gThr...IleSerGlnAlaProAlaArgGluArgProArgLysAsp. 35
      |||||
      57252 CACATCAGTCGGGTGTCTCCAGCGCACCGGTGCGCGGAGAGCATCGG 57203

      36 .....ProLeu.....ArgHisLeuArgThrArg 43
      |||||
      57202 CACACCGGTTACCGAACTCGACCGCACTGGCGGTTTCCCGGTACTCGT 57153

      44 GluLysArgGlyProSerGlyCysSer.....GlyGlyProAsnTh 57
      |||||
      57152 CCAGAATCGGTTCCATCGGTGCGAGTGGAAAGCATGGGAGACCGCGAAC 57103

      57 rValTyrLeuGlnValAlaAlaGlySer.ArgAspSerGlyAlaAla 73
      |||||
      57102 CATTTGG.....TCCTAAGCCCTCGGCG..... 57080

      74 LeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysGlyGluGl 90
      |||||
      57079 ATCGAATTCGGCTCGGAGCTCCAGGAGACATCGCGGTGCGCGGAGAG... 57032

      90 yValcIlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLeuAspA 107
      |||||
      57031 .....CACTACCGACTCGGCGCGGTGATC 57007

      107 sn.IlePheLeuThrArgMetHisTrpSerAsnValGlyLeuSerGI 123

```

```

      57006 GCAGCGATGCTGACCCGATCGCACACCCCGGAAAGCAGCGGTCCGACCGC 56957
      123 yMetIleLeuThrLeuLysGluThrGlyLeuProLysCysValLeuSerG 140
      |||||
      56956 CGCCTCGCTGGTGGCGCGGAGCAT.....TGGCGCGCGGTG 56917

      140 lyProProGlnLeuGluLysTyrLeuGluAlaIleLysIlePheSerGly 156
      |||||
      56916 GG.....CAATGCCTGCATCAGGTG.....GGC 56894

      157 proLeuLysGlyIleGluLeuAlaValArgProHisSerAlaProGluTyr 173
      |||||
      56893 CCGAGCGCAACCAACCGCTGCGCGCTCGCAAGGACACACACCGACCGG 56844

      173 rGluAspGluThrMetThrValTyrGlnIleProIleHisSerGluGlnA 190
      |||||
      56843 CGTGGCGCGCGACCCAACTCGCGGACCGAGTGCCT.....CAGC 56806

      190 rArgGlyLys.....HisGlnProTrpGlnSerProGluArgPro 203
      |||||
      56805 ACCGGCGCGCGCGCGCACGCCCGCACGACCTCAGCAATCCAGGAGCGCGAC 56756

      204 LeuSer.....ArgLeuSerProGluArgSerSerAspSerGI 216
      |||||
      56755 TTGAACCGCGAAGATTCGCCGCTGCGCCACACGCTCGGTGCGAGCATCT 56706

      216 userAsnGluAsnGluProHisLeuProHis..... 226
      |||||
      56705 GCTCATCAGCAGCGACGACACACCCCGAAGCTCGCGCGCGCGAGTGTGCG 56656

      227 .....GlyValSerGlnArgArgGlyVal 234
      |||||
      56655 CTTAGCGCGTTCGAGCGTTCCTCGAAGCGCTCGCGAAGCGCGGGAAGC 56606

      235 ArgaspSerSerLeu.....ValValAlaPheIleCysLys 246
      |||||
      56605 CGCGCACAAACGCTTTCCCATCCCGACGCACTGACTGCTGCGTCCGCGGAAA 56556

      246 s.....LeuHisLeuLysArgGlyAsnPheLeuValL 257
      |||||
      56555 AAACGAATCCGATCGCGCGCACCGGAAATTCGAGACCGCGCTTATCACCCCC 56506

      257 euLysAlaLysGluMetGly.....Leu 264
      |||||
      56505 GCGCGCGTTCGCGCATCGCGAGCGCTGCCAGCTGGACAGCAGTGCCTC 56456

      265 ProValGlyThrAlaAlaIleAlaProIleIleAlaValLysAspGI 281
      |||||
      56455 CCGATCGGCGACCCCA.....CAGCAGG 56433

      281 yLysSerIleThrHis.....GluGlyArgGlu.....IleL 292
      |||||
      56432 CGCGTTCGTCAGCACAGACGAGTGTGCGGAGTGAATACCCCAATCC 56383

      292 euAlaGluGluLeuCysThrProAspProGly.....AlaAla 305
      |||||
      56382 AGCGGAGAAAGCGCGAGGACCCCGGTCCAGCGCGGCTGCAGCGACGCGC 56333

      306 PheValValValGluCysProAspGluSerPheIleGlnProIleCysGI 322
      |||||
      56332 TTGGCGCTGCGCGCTGCGAG.....ACTCCTAGCCGACA 56298

      322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
      |||||
      56297 ACATCCAGCGGATCACCGCTGGATG...CGCAGTACCGCGCGATCCGA. 56252

      339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
      |||||
      56251 .....TTGGTCGATCTCCGC..... 56237

      356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGI 372
      |||||

```



```
||||| ..... ||| |||||:
3397 GGGTCGGGAGAGCTTCTCGAAGT.....CGAGACCGCGCGCA 3439
433 TrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValG1 449
||||| ||| |||||:
3440 ACACACGCGCTGCATCAGGTCTATGTCACCGGGCGAC..... 3478
449 uAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArgArgS 466
||||| |||||:
3479 .....CGGAGGCACGTACACGCTGAACGTCAC 3506
466 erAlaGlnAspGlyProAlaProAlaGluLysArg_SerGlnTyrProG1 482
||||| |||||:
3507 GGACCAACACCGCGCCATCGACTACACGAGCTGCGAGAAATCGCA 3556
482 uIleIle.....PheLeuGlyThrG 489
: |||:
3557 GGGCTGGCAGCGGACATCAACAGCGCTCTGGACGCGCGTCGGAGCGG 3606
489 lySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuValAsn 505
||||| |||||:
3607 GAAACGCTGTCGTGCATCCGCTCTCGTGTCTCCCTGCGTGACACTGAAC 3656
506 IleSer.....ProAspThrSerLeuLeuLeuAspCysG1 517
||||| |||||:
3657 TTCAACCTCAACGCCAAGCGCTCAACGAGCGTGTGTCAC..... 3701
517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
||||| |||||:
3702 .....ACGATCAACAAGCGCGGACGAGCTCTTCGACACGTCGACC 3744
534 rgValLeuGly 537
: |||||:
3745 AACTACTGGG 3755
```

seq_name: /cgn1_7/ptodata/1/ina/5B_COMB.seq:us-09-075-904-1

seq_documentation_block:

; Sequence 1, Application US/09075904

; Patent No. 5994137

; GENERAL INFORMATION:

; APPLICANT: Jacobs, et al.

; TITLE OF INVENTION: L5 SHUTTLE PHASMIDS

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Amster, Rothstein & Ebenstein

; STREET: 90 Park Avenue

; CITY: New York

; STATE: New York

; COUNTRY: U.S.A.

; ZIP: 10016

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette

; COMPUTER: IBM PC Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Word Processor (ASCII)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/075,904

; FILING DATE: May 11, 1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/247,901

; FILING DATE: May 23, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Bogosian, Elizabeth A

; REGISTRATION NUMBER: 39,911

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 697-5995

; TELEFAX: (212) 286-0854 or 286-0082

; TELEX: TWX 710-581-4766

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

;

; </

2268 GGC.....CACGTACTACCGCGGCTGTCGGCCACCAGAAAGTGGGGCGG 2308

87 sGlyGluGlyValGlnArgLeu.....LysVal.AlalArgLeuAspAs 107
|||||:|||||
2309 TGGCGAGGGTCTCGAGATCCTCAACTACGAGGGTCCAGGACCGTCCAGG 2358

95 etGlnGluHisLysLeu.....LysVal.AlalArgLeuAspAs 107
:|||||:|||||
2359 TCTCCGACGAGAAGCTCGCGGAAGCCACAGCGGCCACGAGGTCCAGGCT 2408

107 nilePheLeuThrArgMetHisTrpSerAsnValGlycylLeuSerGlyM 124
|||:|||||:|||||
2409 GAACTTCCGCGCGAGGTTCGGAAGAGTGAGTGGCGGGCTCAGGCCGG 2458

124 etIleLeuThrLeuLysGluThrGly.....LeuProLys..... 135
|||:::||| |||||
2459 CGACAGGAACCTACCAGAGACTGGGAGCTGAATTACCGGTCCTCCGCTCCT 2508

136CysValLeuSerGlyProGlnLeuGluLysTyrLeuGluAl 150
|||:|||||:|||||
2509 TTCTGCTGCCAAGTGGCTTTCAGATCAACGGTCCCGG..... 2547

150 alileLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgp 167
:|||||:|||||
2548ATCGTTAAGGGCGGCAACCGATGTCGAC 2575

167 roHisSerAlaProGluTyrGluAspGluThrMetThrValTyrGlnIle 183
|||||:|||||
2576 CACATCAAGCGGGAAGCACACAGCGGTTCGAATCTGCAGGACGCGCTG 2625

184 ProIleHisSerGluGln.....ArgArgGlyLysHis..... 194
|||:|||||
2626 CCATGCTCTCTACGGCAAGAAATCAGCGCGCGAGGCGTAGCCGACGGC 2675

195GlnProTrpGlnS 199
:|||||
2676 GGGAAGTTAGAGCCGCGAGGAGGACACCCGCAACGCCATCTCTGGGCGT 2725

199 exProGluArgProLeuSerArgLeuSerProGluArgSerSer..... 213
|||:|||||:|||||
2726 CGATAAGCGGCGCAGGTCCCGCTCCACCCAGGAGGTGAACATGTTGGGCAC 2775

213 213

2776 GCGAGGCCCAATCGGAAAAACGAGATGAAGAGCGGGTTCGTCGGAACACCC 2825

213 213

2826 CGGACAGTCCAACCGACACAGATCATGCCCCGTCTGTGTGAGGATCCCC 2875

214 ...AspSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSer 229
|||||:|||||:|||||
2876 GAGATGGCGGATTAAGCCAGCAGCGCGGCCACGACCGACGCTCGTCAAGGA 2925

230 GlnArgArg.GlyValArgAspSerSerLeuValAlaIlePhe..... 243
::|:|||||:|||||
2926 CATGTACGATCGATCAAGCAGTCCGCGAGCGGTCAAGTACTACGACCCGA 2975

244IleCysLysLeuHisLeuLysArgGlyAsnPhelLeuVal 256
|||:|||||
2976 CCGACTGCAGATGGCCGACTCGCCCTCTACACACTTAACACAGGAACCT 3025

257 LeuLysAlaLysGluMetGlyLeuProValcylThr.....AlaAl 270
::|:|||||:|||||
3026 ATCGACCGGAGAACACGCGAACCCGTCGGGCGGATGAAGTCTACTGCTC 3075

270 alleAlaProIleIleAlaValLysAspGlyLysSerIleThrHisG 287
|||||:|||||:|||||
3076 CATCAACACGATGCTCTCCGGCTG.....CTGCTGACCG 3110

287 luGly.....ArgGluIleuAlaGluLeucysThrProProAsp 301
|||||:|||||:|||||
3111 AAGGTGACCGGACCGGCTCCGACTCGAAGTCGAAGCAGACGACCGCTGAC 3160

302 Pro.GlyAlaAlaPheValValGluCysProAspGluSerPheIleG 318
 3161 CCACACAGCGCGGA.....AGGTCGTGGAGCT 3186
 318 InProIleCysGluAsnAlaThrPheGlnAArgTyrGlnGlyLysAlaAsp 334
 3187 GACCGACGTGCTCAACAGCGCTCGCAAGAGCGAGCGCGGGAGCTGAT 3236
 335 AlaProValAlaLeuValHisMetAlaProAlaSerValLeuValAs 351
 3237 GGTCCC.....CGAGGGGTTCAGAGCCG 3262
 351 pSerArgTyrGlnGlnTrrMetGluAArgPheGlyProAspThrGlnHisL 368
 3263 CTCCTCCCTACACAGCGCTC.....CCCTCGGGGTAGACATCGAA... 3302
 368 euValLeuAsnGluAsnCysAlaSerValHisAsnLeuAArgSerHisLys 384
 3303AGGAACCAACATG 3314
 385 IleGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuTh 401
 3315 GCGGACCTCGGC...AACCACTCGACCTCGAGATGCTCTCCCTGGTCAC 3361
 401 r.....SerPheAArgCysLysLysGluGlyProThrLeuSerValProM 416
 3362 AGCGCGGAGCTTCGCTGG.....ACCATCGATTACCCGT 3396
 416 etValGlnGlyCysLeuLeuLysTyrGlnLeuAArgProAArgArgGlu 432
 3397 GGGGTCCGGAGAGCTGTTCCTCGAACT.....CGAGACCGCGGCGGA 3439
 433 TrpGlnAArgAspAlaIleIleThrCysAsnProGluGlnPheIleValG 449
 3440 ACAACACGGCTGCATCAGTCTATGTACCGGGGCGAC..... 3478
 449 uAlaLeuGlnLeuProAsnPheGlnSerValGlnGluTyrArgArgS 466
 3479CGAGCAGCTACACGCTGAACGCTCAAC 3506
 466 erAlaGlnAspGlyProAlaProAlaGluLysArg.SerGlnTyrProG 482
 3507 GGCACCAACACCCGGGCCATCGATTACAACGAGCTGTCCGGAATCCGGA 3556
 482 uIleIle.....PheLeuGlyThrG 489
 3557 GGGGTGGCAGGCNACATCCAAGAGCTCTGGAGCAGCGCTCGAGCCG 3606
 489 lySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuValAsn 505
 3607 GAAACGCTGTGTGCATCCGGTCTCGTGTTCCTGCGTGGACACATCAAC 3656
 506 IleSer.....ProAspThrSerLeuLeuLeuAspCysG 517
 3657 TTCACACTCAACGCCAGCAACCGCTCACCAGCAGATTGTGTCAAC..... 3701
 517 yGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspA 534
 3702ACGATCAACAGCGCGGACAGCACTTCTTCGACACGTTGCACC 3744
 534 rgValLeuGly 537
 3745 AACTACTGGG 3755

seq_name: /cgn1_7/ptodata/1/ina/5A_COMB.seq:US-07-642-734C-3

```
seq_documentation_block:
; Sequence 3, Application US/07642734C
; Patent No. 5824513
;
; GENERAL INFORMATION:
;
; APPLICANT: Katz, L
;
; APPLICANT: Donadio, S
```

APPLICANT: Mcalpine, J B
TITLE OF INVENTION: Recombinant DNA Method for Producing
Thrombolytic Enzymes
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Edward H. Gorman
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dancckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:

NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature

```

: LOCATION: 16768..17721
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: acyltransferase domain of module 6"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 18379..18921
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: beta-ketoreductase domain of module 6"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 19149..19398
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: acyl carrier domain of module 6"
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 19492..20235
: OTHER INFORMATION: /function= "approximate span of
: OTHER INFORMATION: thioesterase domain of module 6"
: US-07-642-734C-3

```

```

alignment_scores:
  Quality: 115.00      Length: 751
  Ratio: 0.389        Gaps: 40
  Percent Similarity: 39.414      Percent Identity: 21.039

alignment_block:
  US-09-4347-382-2 x US-07-6442-734C-3/rev ..

```

Align seg 1/1 to reverse of: US-07-642-734C-3 from: 1 to: 20235

10	SerAlaIaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAla	Pr 26
	::: :::	
9943	TCGGGGGAAGCCGCGCTCTCCGGAGGCGAGCC 9909
26	oAlaArg.....ArgGluArgProArgLysAspProLeuArgHis	L 40
	::: ::: :::	
9908	AGCGGTTGACGGACTGTTCCGCGCGCGGTACACGCCCTGCCCGGGGCC	9859
40	eu.....ArgThrArg...GluLysArgGlyProSerGly	50
	::: ::: :::	:::
9858	GGCTAGCACCCAGCGCGGACGAGAAGACACCAAGAAGCTCAGACCGG	9809
51	CysSer.....GlyGly_ProAsnThrValTyrLeu.....	60
	::: :::	
9808	TGTTCCGGCTCAGCTGTCGAGGTTCCACGCCGCTCGACCTTGCGCGCG	9759
61GlnValValAlaAlaGlySerArgAspSerGly	71
	::: ::: ::: :::	
9758	AGCACCTGCTCCACCGCGCGCTGCTCGATGGAGTACACAGCCGTCGGC	9709
72	AlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysG	88
	::: ::: ::: :::	
9708CAGCACACCCCGCGGTGCACACACCGG	9680
88	yGluGlyValGlnArgLeuMetGlnGluHisLysLeuLysValAlaAargL	105
	::: ::: ::: :::	
9679	TCAGCGGCGGGGAGCCCGTCCAGCAGCGCGAAGACGGCGTCCGCG	9634
105	euAspAsnIlePheLeuThrArgMethIstirpSerAsnValGluGlyLeu	121
	::: ::: ::: :::	
9633GTGCGCGGGTGC	9622
122	Ser.....GlyMetIleLeuThrLeuLys.....	129
	::: ::: ::: :::	
9621	GCAAGCGCGGATCTCCGCGACGCGCCAAAGTCTCGATCCGCGCGCA	9572
130GluThrGlyL	133
	::: ::: ::: :::	
9571	GCTGCTGGAAACCGCGCGCTCCGCGCGCGCTGCTGACCAACGACGAGG	9522
133	euProLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGlu	149
	::: ::: ::: :::	

[illegible]

370 LeuAsnGluAasnCysAlaSerValHisAsnLeuArgSerHisLysIleGI 386
:::|||||::: |||||::: ||| :::: |||||
8628 TCGAGGAGGACGTCCGGAAATGTACTTCGGTGC CGCAC..... 8587
386 nThrGlnLeuAsnLeuIleHisProAspIlePhe.....ProLeuLeu 401
:::|||||::: |||
8586GTCCACCCGGACCTGCCGCCGCCGAGCGGTGC 8553
401 hrSerPheArgCysLysLysGlyGluGlyProThrLeuSerValPrometVal 417
::: ||| ||||| :::: |||||
8552 TCGACGTCCGGTGCGGGCTCGAAGGCCACGAGGTGATGCCCGTCACC 8503
418 Gln.....GlyGluCysLeuLeuLysTyrGlnLeuArg...Pr 429
::: ||| ||| :::: |||
8502 CGCGACGAGACGTTCGGTCCCTCGCGGAAGCGTCAGCGCGAAGTCCGA 8453
429 oArgArgGlnTrpGlnArgAspAlaIleIleThrCysAsnProGluGluP 446
||||||| :::: |||||
8452 GCGCGGAGAGCGCGGCACGAAGAAGCTGTCGCCGCCGCCAACGAGCTGC 8403
446 heIleVal...GluAlaLeuGlnLeuProAsnPheGlnGlnSerValGln 461
::: ||| :::: ||| |||||::: |||||
8402 GGTTGTCCGGNACCGCAGGACGATCCCAGGTCCCGCTCGGTGTGCAG 8353
462 GluTyrArg.....ArgSerIadGlnAspGlyPr 471
::: ||||| :::: |||||
8352 GAGCACAACCGTTCGGCGCATCTCGGCTCGCACACGCGATCAGCGCCC 8303
471 O..... 471
8302 ACATCCGCGCGCGCGCGGTTTCGGGAACGCTGTCGTCTCGGACACGGTG 8253
471 471
8252 ACCGCGCGGACGTGGCATGACACAGCGTGCGCCCGCAGCTCCTCTGT 8203
472 ..AlaProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGI 487
||||| :::: |||||
8202 TTCAGACCAAGCGCGCACGACGCGCGCGGCCACAGACGCCGTGGCCGG 8153
487 yThrGlySerAlaIlePrometLysIleArgAsnValSerAlaThrLeuV 504
||| :::: |||||
8152 CCGCGCGCG.. 8143
504 alGsnIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyThr 520
||| :::
8142GGGTCGTGCGCC 8131
521 PheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgVal..... 535
||||| |||||
8130 TTCGGG.....TTCGTATCGCGAACAG...GACCGCGTCGGGTT 8096
536LeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaA 550
||| :::: |||||
8095 CGGTTACCGGACCGCAGGCGGTTCGAGTCCGAGGATCGGCCGCG 8046
550 sPHisHistrGlyLeu.....ProSerIleLeuLeuGln 561
||||| |||||
8045 ACCACACACCGGTTCGAGCGCGGTTCGGTTCGCGTCCCGACGTCACAGC 7996
562 ArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeu 577
||||| |||||
7995 GTGACGTCGCGCTCCGACCGCGGTTCGGGTGTCGTCACATTCTC 7948

seq_name: /cgn1_7/ptodata/1/ina/6_COMB.seq:US-08-439-009A-3

seq_documentation_block:

Seq=documentation_LOCK.
; Sequence 3, Application US/08439009A

; Patent No. 6004787

GENERAL INFORMATION:

APPLICANT: Donadio, S

APPLICANT: Katz, L

110

```

APPLICANT: McAlpine, J B
TITLE OF INVENTION: Method of Directing Biosynthesis of
TITLE OF INVENTION: Specific Polyketides
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Steven F. Weinstock
STREET: Abbott Laboratories D377/AP6D-2 One Abbott
CITY: Park Rd
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,009A
FILING DATE: 11-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Casuto, Dianne
REGISTRATION NUMBER: 40,943
REFERENCE/DOCKET NUMBER: 4952-US.D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 847-938-3137
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start= 19
OTHER INFORMATION: /function= "gene eryA"
OTHER INFORMATION: /product= "eryA ORF2 encoding modu
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of modu
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of modu
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature

```

```

LOCATION: 4471..10722
OTHER INFORMATION: /function= "approximate span of
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..5847
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPsynthase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function= "approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start= 10723
OTHER INFORMATION: /function= "gene -eryA"
OTHER INFORMATION: /product= "orf3 encoding modules 5 & 6"
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function= "approximatr span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoacylACPsynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721

```

```

OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19149..19398
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: acyl carrier domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19492..20235
OTHER INFORMATION: /function= "approximate span of
OTHER INFORMATION: thioesterase domain of module 6"
US-08-439-009A-3

alignment_scores:
Quality: 115.00 Length: 751
Ratio: 0.389 Gaps: 40
Percent Similarity: 39.414 Percent Identity: 21.039

alignment_block:
US-09-434-382-2 x US-08-439-009A-3/rev ..
Align seg 1/1 to reverse of: US-08-439-009A-3 from: 1 to: 20235

10 SerAlaAlaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAlaPr 26
||| |||||: ||| |||||: ||
9943 TCGCGGGAAGCGCGCGCTCCCGAGGGCAG.....CC 9909

26 oAlaArg.....ArgGluArgProArgLysAspProLeuArgHisL 40
||||| |||||: |||||: |||||
9908 ACCGCGTTCAGGACTGCTTCGCGCGCGGTACACGCCCTCCCGCGGCC 9859

40 eu.....ArgThrArg...GluLysArgGlyProSerGly 50
||||| |||||: |||||: |||||
9858 GCGTAGCACCGACGCGCGGACGACAGCAGCAGCAAGAGTCAGACCGG 9809

51 CysSer.....GlyGly.ProAsnThrValTyrLeu..... 60
||||| |||||: |||||: |||||
9808 TGTTCGCGTCAGCTCGTCCAGGTTCCACGCGCGCTCGACCTTGGCGGC 9759

61 .....GlnValValAlaAlaGlySerArgAspSerGly 71
||||| |||||: |||||: |||||
9758 AGCAGCTGCTCCACGCGCGGCTCGTGGAGTGACCGCCCGCTCGGC 9709

72 AlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnCysgl 88
||||| |||||: |||||: |||||
9708 .....CAGCACACCCCGCGCGTCCAGCACACCGG 9680

88 yGluGlyValGlnArgLeuMetGlnGlyHisLysLeuLysValAlaArgL 105
||||| |||||: |||||: |||||
9679 TCAGCGCGCGGGCAGCGCTCCAGCAGCGCGGAAAGCGCGTCGCGC... 9634

105 euAspAsnIlePheLeuThrArgMetHisTrpSerAsnValGlyGlyLeu 121
||||| |||||: |||||: |||||
9633 .....GTCGCGCGGTGTC 9622

122 Ser.....GlyMetIleLeuThrLeuLys..... 129
||| |||||: |||||: |||||
9621 GCAAGCCGCGATCTCCGCGGAGCGCCCAAGTCCTCGATCTCCGCGCGCA 9572

130 .....GluThrGlyL 133
||| |||||: |||||: |||||
9571 GCTCGTCGGAACCCGCGCGTCCGCGCGCGCTGCTGACACGACGAGG 9522

133 euProLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGlu 149
||||| |||||: |||||: |||||
9521 TGCCGCGACCGCTGCTCGGTGACGAGGTGGCGCGGACGCGCTCCGAG 9472

```

150 AlaIleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValAr 166
9471 CGT.....GCCGGTCCACCCCGTGTATCAGACACCGTCCCGTCCG 9434
166 gProHisSerAlaProGluTyrgluAspGluThrMetThrValTyrglnI 183
9433 GTTCACCCGGCGGGCTGGTGCAGCAGCGTTCGCCGACGTG.....CCTG 9387
183 leProIleHisSerGluGlnArgGlyLys..... 193
9386 CCCCGGCTCATGTGCTCAACCGCGCGCGGATCCAGCTCCACACG 9337
194HisGlnProTrpGln_Ser.....ProGluArgProLeuSer. 205
9336 CGATACCGGGAGCGCGTGCAGTCCCGCGCGCGAGCGACGACCT 9287
206ArgLeuSerProGluArgSerSerAspSerGlnSerAsnGluAsn 220
9286 CGCGCAGGATCTCCCGCGCGCGGTCTACCCCGCTTCGCCGAGGTGGAAC 9237
221 GluProHisLeuProHis...GlyValSerGlnArgGlyVal..... 234
9236 GGGGGGTATCGGCCCGGGAAGTCCCGCGGTCCCGCAGGTCTGTGCC 9187
235ArgAspS 237
9186 CATCTCGACGAAGACCGCTCTCCGCGAGCAGATCCCGCGGACTCTCGA 9137
237 erSerLeuValValAlaPheIleCys..LysLeuHisLeuLysArgGlyA 253
9136 GCAGTTCCTCCGGTGGAGCGGTGAGCACACCGTCCACCGCGCGGCTCCG 9087
253 snPheLeuValLeuLysAlaLysGluMetGlyLeuProValGlyThrAla 269
9086 GTCCGCTCCCGGAACCTCCGGGGAACCGGTCTCCCGGAGAGCGAT 9037
270 AlaIleAla.....ProIleIleAlaAlaVa 278
9036 GTCTCGTCTCGTGCAGACGAGCGCGCGCAGCGTCCCGTGTGGCGGGC 8987
278 llysaspGlyLysSerIle.....ThrHisGluGlyA 289
8986 CGCGGTGGCCACACCTCCGCGCGCGCTCCGCGCGCGCGCGCGCGC 8937
289 rgGluIleLeuAlaGluGluLeuCysThrProProAspProGlyAlaAla 305
8936 GCAT.....GCCGACCGCGCTGC..... 8917
306 PheValValValcLysCysProAspGluSerPheIleGlnProIleCysgl 322
8916CGTTCGCTGGATGAGCAGCAGCTACCCCGCGCGCAGCC 8879
322 uasnAlaThrPheGlnArgTyr.....GlnG 331
8878 CGCGAGATCGTGTGCAGCGGTAAATGCGCGCTGGTGTAGCGATGGCGACG 8829
331 lylsAlaAspAlaProValAlaLeuValValHisMetAlaProAlaSer 347
8828 GCGCGGCTGGCGTGTCTCAGCGCTCCGGCAGCTCGGCGAGGAGCGG 8779
348 ValLeuValAsp...SerArgTyrGlnGlnTrpMetGluArg..... 360
8778 GTGATCGGTGACCGCGATCGCGCGAAGCTCCCTTGAACAGCGCGCGCA 8729
360 360
8728 CCGGTCTCCCGCGCGAGAGCGTCCACGTCCGCTCCGCGCGCGTACG 8679
361PheGlyProAspThrGln.....HisLeuVal 369
8678 ACACCGCGCGCTCGGTGCCATGTCCGCTTCTGCGGATACATCGCGAG 8629

370 LeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGl 386
8628 TCGAGGAGGAGCGTCCGGAAGTTGACTCGGTGGCGCGCAC..... 8587
386 nThrGlnLeuAsnLeuIleHisProAspIlePhe.....ProLeuLeuT 401
8586GTCCACCCGAGCCTCGCCCGCGCGGAGCGGTGC 8553
401 hrSerPheArgCysLysLysGluGlyProThrLeuSerValProMetVal 417
8552 TCGACGTGGTGGCGGTTCGAAGGCCACGAGTTCGATCGCCCGTCACC 8503
418 Gln.....GlyGluCysLeuLeuLysTyrGlnLeuArg...Pr 429
8502 CGGCACGAGACGTGGGTGCTCCCGGAAGCGTCAGCGGAAGTCCGA 8453
429 oArgArgGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluP 446
8452 GGGCGAGAGCGCGCACGAAGACGTCTGCGCGCGCACACGCGAGTGC 8403
446 heIleVal...GluAlaLeuLeuProAsnPheGlnGlnSerValGln 461
8402 GGGTTCTCGGAACCGCAGCAGCATCCAGGTTCGCGTGGTGTGAG 8353
462 GluTyrArg.....ArgSerAlaGlnAspGlyPr 471
8352 GAGCAGAAACCGGTCCCGCGACTCGGCCTCGCACACGCGGATCACGCC 8303
471 O..... 471
8302 ACATCGCGCGCGCGGTTCGGGAACGCTGCTGCTGGGACACGGTG 8253
471 471
8252 ACCGCCCGGACGTGGCGATGACCGCTCGCCCGCGGACGTCTCTCTG 8203
472 ...AlaProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGl 487
8202 TTCGAGCAGCGCGCAGCAGCGCGCGCGCCAGAGACGCCGTGGCGG 8153
487 yThrGlySerAlaIleProMetLysIleArgAsnValSerAlaThrLeuV 504
8152 CGCGCGCGCG..... 8143
504 aIAsnIleSerProAspThrSerLeuLeuLeuAspCysGlyGluGlyThr 520
8142GGGTGCTGCGCC 8131
521 PheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArgVal..... 535
8130 TTCGGG.....TTCGTAGCGGACAG...GACCGCTCGGGTT 8096
536LeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaA 550
8095 CGGGTTTACCGGACCGCAGCGCGTTCGAGGTCCGAGGATCGCGCGCG 8046
550 sphiHisThrGlyLeu.....ProSerIleLeuLeuGln 561
8045 ACCACAGCAGCGTTCGAGCGCGGTTCGGTAGCCGCTCCCGACGACAG 7996
562 ArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeu 577
7995 GTCCAGCTCGCCCTCGCCACCGCGCGGTTCGGGTGCTGCCACTTCTC 7948
seq_name: /cgnl_7/ptodata/1/ina/5A_COMB.seq:US-08-093-453B-1
seq_documentation_block:
; Sequence 1, Application US/08093453B
; Patent No. 5439814
; GENERAL INFORMATION:
; APPLICANT: Frey, Teryl K.
; APPLICANT: Dominguez, Geraldina
; APPLICANT: Wang, Chin_Yen


```
4913 GCAUC.....CCGAGGCCAAGACGCGCCCUUCUGCCC 4950
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
322 uAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaL 339
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
4951 CGCACUCUGGAGGAGCUCUCUCCGCGGCGGCCGCCACCC..... 4993
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
339 euValValHisMetAlaProAlaSerValLeuValAspSerArgTyrGln 355
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
4994 .....CAUACGCGGAGCCUCCACCGGUGUGACGAGGGCGAACAGAA 5035
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
356 GlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnG 372
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5036 .....GUGGGUACAGGCGCAUCUCGCGUACCG.....CUCAACAA 5073
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
372 uAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGln 389
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5074 GAU.....CACACCGAGA 5087
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
389 euAsnLeuIleHisProAspIlePheProLeuLeuThrSer..... 402
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5088 UG.....CCCGAAGGAACGGGUUCACAGUGCGGUUGCGGC 5125
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
403 .....PheArgCysLysLysGluGlyProThrLeuSerValProMe 416
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5126 GUGCGCGCUACCGCGGCGGAGGUGGUGGACCCUCCGACUCUGUGU 5175
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
416 tValGlnGlyGluCysLeuLeuLysTyrGlnLeuArgProArgArg.... 431
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5176 GCGCCCGCAGCACCGCGCCUUCUCCGAGAUCCACCCCGCGGUGCA 5225
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
432 .....GluTrpGlnArgAspAlaIleIleThrCysAsn 442
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5226 CUGCUGGGUGCGCCAGAGGUGGCGAUG..... 5254
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
443 ProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGlnInSe 459
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5255 .....ACGUACUUGCGGGAAGCGAUCGACCUACUCAGUCUACACGCA 5298
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
459 rValGlnGluTyrArgArgSerAlaGlnAspGlyProAlaProAlaGlu 476
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5299 GAUGGCGUGCGCGCGGAGCUCACCCAGCCUACGCG..... 5338
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
476 ysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySerAlaIle 492
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5339 .....CGCGCUAUCCUGAG...AUCUUCGCGGCAUGUGUACCGCCAG 5380
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
493 ProMetLysIle...ArgAsnValSerAlaThrLeu.....Va 504
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5381 ACCUUGAGCGUCCCGCCUUCUCCAAAGCCACCUUGAAGUGGUGUACGCG 5430
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
504 lAsnIleSerPro...AspThrSerLeuLeuLeuAspCys.....GlyG 518
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5431 CGCCUUGCGCCCGGACACCGAG.....GACUCCCGCGCGCUC 5471
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
518 luGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspArg 534
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5472 AGGGAAAGCGCGCCUUGAUCGCGGCGUGGCGCAAGGAGGUGGUUCAG 5521
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
535 valLeuGlyThrLeuAlaValPheValSerHisLeuHisAlaAspH 551
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5522 GUUAUG.....UCCCGCA 5535
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
551 shiThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlas 568
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5536 UUUCCCGCGAUCCAGAAUAUCAUG..... 5563
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
568 exLeuGlyLysProLeuHisProLeuValValAlaProAsn..... 582
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5564 .....CGCGCUUGCGCGCCAAUCCUUGGCGCGCGCGCAUACG 5605
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
583 .....GlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGln 597
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
```

```
5506 GACCCCGAGUGUGAUGCGUGGUGGAG..... 5632
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
597 uValLeuHisHisIleSerMetIleProAlaLysCysLeuGlnGluGlyA 614
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5633 ..... 5633
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
614 laCluIleSerSerProAlaValAlaGluArgLeuIleSerSerLeuLeu 630
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5634 CCAUACACCCACCAACGCAUCGAG..... 5659
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
631 ThrCysAspLeuGluGluPhe.....GlnThrCysLeuValArgHi 644
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5660 .....GUCGACUUCACUGAGUUGGACAUCAACACCCUCCUACUGGGA 5706
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
644 sCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLys 661
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5707 CGUCGAGCUGAGAUUAGCGCCGUCUCUUG..... 5737
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
661 alValTyrSerGlyAspThrMetProCysGlu.....AlaLeuVal 674
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5738 .....GGCCUCCUUGCGCGAAGACUACCGCGGCGUC 5770
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
675 ArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGlu 691
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5771 CGCGCGCGGAGUACUGACCCUG.....CGCGAAUC 5802
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
691 pGlyLeuGluGluAlaValGluLysThr..... 701
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5803 GGCUCUACUGAGACCGGUGCGGCGCACAGCGGCGGCGCCGCGCAGC 5852
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
702 .....HisSerThrThrSerGlnAlaIleSerValGlyMetArgMet... 715
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5853 UGCUGCACAAACACACC...GUGCGCAUGUGCAUGGCGCAUGCGCAUGGUC 5899
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
715 ..... 715
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5900 CCAAAAGCGUGCGUGGCGCGCGGAUUUCCAGGGUGACGAUUGUCAU 5949
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
716 .....AsnAlaG 718
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
5950 CUUCCUCCGAGGCGCGGCGGCGGCGGCAUCUAGUGACCCCGCGG 5999
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
718 luPheIleMetLeuAsn...HisPheSerGlnArgTyrAlaLysValPro 733
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
6000 AGGUGGCUUGUUGCUUCCACUCCGCGGUGAAGCAGCUGAGCACCCCU 6049
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
734 LeuPheSerProAsnPheSerGluLysValGlyValAlaPheAspHisMe 750
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
6050 .....ACCCCGAGUUCUGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6088
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
750 tLysValCysPheGlyAspPheProThrMetProLysLeuIleProL 767
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
6089 .....CUUCCUAGAU.....GUCAUGCACCGGCGCAUCAGGUGC 6125
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
767 euLysAlaLeuPheAlaGlyAspIleGluGluMetGluGluArgArgGlu 783
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
6126 UUUUGCGCGUUCGACCCAGCAGCUGCUUGAAGAACAACAGGUGGCCUC 6175
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
784 LysArgGluLeuArgGlnValArgAlaAlaLeu 794
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
6176 CUCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6208
      :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
```

seq_name: /cgml_7/ptodata/1/ina/6_COMB.seq:US-08-938-291A-1

seq_documentation_block:

; Sequence 1, Application US/08938291A

; Patent No. 6117673

; GENERAL INFORMATION:

; APPLICANT: Lev, Sima

; APPLICANT: Plovman, Gregory D.

; APPLICANT: Schlessinger, Joseph

; TITLE OF INVENTION: RDB PROTEINS AND RELATED

; PRODUCTS AND METHODS

```

: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Lyon & Lyon
: STREET: 633 West Fifth Street
: STREET: Suite 4700
: CITY: Los Angeles
: STATE: California
: COUNTRY: U.S.A.
: ZIP: 90071-2066
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
: MEDIUM TYPE: storage
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: IBM P.C. DOS 5.0
: SOFTWARE: FastSeq
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/938,291A
: FILING DATE: September 26, 1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/027,337
: FILING DATE: October 11, 1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Warburg, Richard J.
: REGISTRATION NUMBER: 32,327
: REFERENCE/DOCKET NUMBER: 228/172
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 489-1600
: TELEFAX: (213) 955-0440
: TELEX: 67-3510
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3109 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-938-291A-1

alignment_scores:
    Quality: 114.00      Length: 614
    Ratio: 0.460        Gaps: 31
    Percent Similarity: 40.391    Percent Identity: 20.195

alignment_block:
US-09-434-382-2 x US-08-938-291A-1
..
Align seg 1/1 to: US-08-938-291A-1 from: 1 to: 3109

9  ArgSerAlaAlaGlyArgThrMetSerGlnGlyArgThrIleSerGlnAl 25
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
468 AGGGAAGCATCGAGATCCACGAAGACAGCGAGGAAGCTGCCCGCAGCGC 517

25  aPronAlaArg..... 29
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
518 TCCTGCAAGACACATGTCCTCTGCTGCTGTCATGGGGAAACATCCT 567

30  .....GluArgProArgLysAspProLeuArg.HisIle 40
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
568 GGACACGGGTCCGGGGACCGTCTCTGCAAGCAGCCGACATCCACCT 617

40  uArgThrArgGluLysArgGly.....ProSerGlyCysSerGly 54
|:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
618 TCAGCTCGGTGCGAGAAGGTCACACGAGCCCATTTCCCTGCTGCCCTG 667

54  lyProAsnThrValTyrLeuGlnValAla..AlaGlySerArgAspSe 70
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
668 GGCCACATCCTCATCAAGATTCGTCCTGCCCTGCTGCTGCCATCTCTGAG... 715

70  rGlyAlaAlaLeuTyrValPheSerGluPheAsnArgTyrLeuPheAsnC 87
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
716 .....GCTTCTCGCTGTCTCTCACCTGACCCCTACAGCCACGAT. 757

```

```

87  ysGlyGluGlyValGlnArgLeuMetGlnGlnHis..... 98
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
758 .....GAGGGCTGCCTCAGCAGCAGCAGCAGCAGCTCCCTCTGCGCGCC 802

99  .....LysLeu 100

803  CTTCCCTGTTGGCCATCTCTCCCGCAGTACCAGGATGCTGTGCGCCAC 852

100  ulysValAlaArgLeuAspAsnIlePheLeuThrArgMetHisTrpSerA 117
:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
853  CGTCATCGAGCGAGCCAAACAGGCTCTACAGAGATTCCTGAAGTCCCTCTG 902

117  snValGlyGlyLeuSerGlyMetIleLeuThrLeuLysGluThrGlyLeu 133
:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
903  ATGGGATTGGCTTCACTGGGAGGTGTCTCATCGGGAC..... 943

134  ProLysCysValLeuSerGlyProProGlnLeuGluLysTyrLeuGluAl 150
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
944  .....TGTGTGGG....GGGCTCTCTGGCTTCGATGCCATCTG..... 978

150  aIleLysIlePheSerGlyProLeuLysGlyIleGluLeuAlaValArgP 167
:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
979  .....CTACAGTCGGGGCCCTCAGGGA.....CAGC 1006

167  ro...HisSerAlaProGluTyrGluAspGluThrMetThrValTyrGln 182
||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1007  CTTGCCAGCAGCAGCGGAAGGGAGCATCAG.....CAG 1041

183  IleProIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGln.S 199
|||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1042  CACCCAGGACACCCAGTCGCGGTGGAGGAAGATTGCAGCCTGGCCAGCA 1091

199  er..... 199
||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1092  GCAGCGTCTCAGCAAAAGCAACATTGACATCTCCAGTGGGTGGAGGAT 1141

200  .....ProGluArgProLeuSerArgLeuSerProGluArgSerSerAs 214
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1142  GAGGAGCCCAAGAGCGCTTG.....CCGCGGAAACAGAGCGA 1179

214  pSerGluSerAsnGluAsnGluProHisLeuProHisGlyValSerGlnA 231
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1180  CTCCTCCACCTATGACTCGAGGCGCATCCCCAGCACCAT..... 1219

231  rGArgGlyValArgAspSerSerLeuValAlaPheIleCysLysLeu 247
|||||:|||||:|  |||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1220  .....GCCTTCTCTCAAGCATC 1237

248  His..... 248
|||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1238  CACTCCAGCGTGCTAAAGGATGAGTCTCAGACCCCGCGGCTGGGGGCC 1287

249  .....LeuLysArgGlyAsnPheLeuValLeuLysA 259
|||  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1288  GCAGTCCCTGAGTTCAGCCTGGCGCGCTTTGACTTCGATGTGTCCGACT 1337

259  laLysGluMetGlyLeuProValGly...ThrAlaAlaIleAlaProIle 274
:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1338  TCTTCCTCTTCGGCTCGCCACTGGCCCTGGCTGGCCCTCGGGAGGAGC 1387

275  IleAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGluI 291
:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1388  GTGCTGCTGGCTGGAGCGGTTCACAGTGCCTCTCTGCTGCAGCCAGGT 1437

291  eLeuAlaGluLeuCysThrProProAspProGlyAlaAlaPheValV 308
:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1438  CTACAGCTTCTTCATTCGCGACGCCCTCTGCTCAGCGGTTCGAGCCAC 1487

308  alValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAla 324
:|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
1488  TGCTGGAG....CCCAAGTTCACCTGGTGGCGGCTGTC..... 1522

325  ThrPheGlnArgTyrGln..... 330

```

```

1523  ACCTGTCCTCGTACACAGAGTTCACACTGGCGGATGGCGAGTCCCTCTCT 1572
331  .....
1573  CTTGCTGATGCCCTACACACACAGCCCTCTTCTTCTGGAGGCGAGCT 1622
332  ysAlaAspAlaProValAlaLeuValValHisMetAlaProAlaSerVal 348
1623  CCGGACAGCGCCGACCTCTG.....GATGCCCTCGCTCGGCC 1663
349  LeuValAspSerArgTyrGlnGlnTrpMetGluArgPheGlyProAspTh 365
1664  CTTACGGCTCGAGTTCAGCGCCAGCAGGAGGATGAGGAGGGAG 1713
365  rGlnHisLeuValLeuAsnGlnCysAlaSerValHisAsnLeuArgS 382
1714  CTCCACAGCAGAGCTCGGATCTCGACAGCATGCCACCGTGGGTG 1763
382  erHisLysIle.....GlnThrGlnLeuAsn 390
1764  CTTCCCGCATACAGCAAGTGGTGGGGAAGCAAGAGATCGACTATGCC 1813
391  LeuIleHisProAspIlePheProLeuLeuThrSerPheArgCysLys 407
1814  CTGTACTGCCCTGATGTC.....CTCAGCGCTTC..... 1843
407  sGluGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeu 424
1844  .....CCACCGTGGCCCTGCC..... 1861
424  ySTyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaIleThr 440
1862  .....CACCTCTTCCAGCGCAGTACTTGGAGTCCACAGACGTGGTGGC 1906
441  CysAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPhe 457
1907  .....TTCATCTGAGA.....CAGGTAATGCGCTATCA 1935
457  nGln...SerValGlnGluTyrArgArgSerAlaGlnAspGlyProAla 473
1936  GAGCGTGAACATCAAGAAAGCGCCCGCTGGACCTCGCAGCACTAGTC 1985
473  roAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThr 489
1986  CTGCCAACCCCGGAGAAGTGG.....CTTCGT 2014
490  SerAlaIleProMetLysIleArgAsnValSerAla 501
2015  AAGCGGACTCAGCTCAAGCTGAGGATGTCACGGCT 2050

```

seq_name: /cgnl_7/ptodata/1/ina/PTCUS_COMB.seq:PCT-US95-16980-1

```

seq_documentation_block:
; Sequence 1, Application PC/TUS9516980
; GENERAL INFORMATION:
; APPLICANT: Brigham & Women's Hospital
; TITLE OF INVENTION: CONTROLLING TRAF-MEDIATED SIGNALS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street Suite 3100
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16980

```

```

; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/367,540
; FILING DATE: 30-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05311/014W01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)542-5070
; TELEFAX: (617)542-8906
; TELEX: 100254
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 151..1854
; PCT-US95-16980-1

```

```

alignment_scores:
  Quality: 113.50      Length: 672
  Ratio: 0.425        Gaps: 38
  Percent Similarity: 39.732  Percent Identity: 21.726

```

alignment_block:

US-09-434-382-2 x PCT-US95-16980-1/rev ..

Align seg 1/1 to reverse of: PCT-US95-16980-1 from: 1 to: 2359

```

166 ArgProHisSerAlaProGluTyrGluAspGluThrMetThrValTyrG 182
186 CTCTCCCTCTTCTGCC.....GCTTCTCTCCGCTCTCACA 1955
182 nIleProIleHisSerGluGlnArgArgGlyLysHisGlnProTrpGln 199
1954 AGGTCTCTTTCTGAGCGCGAGGACCTCAGTGAAGACAGACGCCGTTCA 1905
199 erProGlu.....ArgProLeuSer 205
1904 CCCCCAGAGGAGTTCCTTCTGCTGCTAAATCCACCTCCCGAGCTACTT 1855
206 ArgLeuSerProGluArgSerSerSerGluSerAsnGluAsn..... 220
1854 GGGATCGGCAGATCCGGAAGTATCCACTATGACTTTAATAAAAAATTG 1805
220 ..... 220
1804 CATCTTTAATATATGTCCTTCTAGAACAGTTTGGGCCCAAGACT 1755
221 .....GluProHisLeuProHisGlyValSerGlnArgArg 232
1754 GGGCAGCCAGAGCGGATATCATCTCTCCAGCGGCTTCTTGAAGCTGCT 1705
233 GlyValArgAspSerSerLeuValAlaPheIleCys..... 245
1704 GCTGTTGGG.....GTCGGGCTTGAATGCATCTCCCAAAAT 1670
246 .....LysLeuHisLeuLysArgGlyAsnPheLeuVal 257
1669 GAGCTCGAGAGGACCCCTGATCATCATCATGAGTGTCTCTTCTGCTTA 1620
257 euLysAlaLysGluMetGlyLeu..... 264
1619 AAGGGCAGGAGGAGCGGATCATATTCTCCAGCATGATGACAAAAA 1570
265 .....ProVal..... 266

```

1569 CAGCGACAAGTGGCGTCCCTTCCCATCCCGTCCCGTTCCAGGTAGACCC 1520
267 GlyThrAlaAlaIleAlaProIleIleAla.....AlaVal 279
||||| ||||| :::: :::: |||||
1519 TGGCACACATCTTATAGCCAAAGTAAACCAGTGTAGAAGGCTGCCTGTAA 1470
279 ysAspGlyLysSerIleThrHisGluGlyArgGluIleLeuAlaGluGlu 295
|| ||| ::::: ||| ::::: |||
1469 AG..GGACAGGGTCTTCCCATCAGCGC..... 1444
296 LeuCysThrProProAspProGlyAlaAlaPheValValValGluCysPr 312
|| ||| ::::: ||| ::::: |||
1443 CTCCTGCTTCGCCG.....CTTGACTCGCAATCTCC 1409
312 oAspGlu...SerPheIleGlnProIleCysGluAsnAlaThrPheGlnA 328
||||| ||||| ::::: |||
1408 AGATGAGCACTCCATTGTAGCTGGCGGTCTCCAGCACTGGAACGGCAGG 1359
328 rgTyGlnGlyLysAlaAspAlaProValAlaLeuValHisMetAla 344
::: ||| ||||| |||||
1358 TCCATGTCGGCTAGCGGAT.....GTCTGTCAC..... 1330
345 ProAlaSerValLeuValAspSerArgTyrGlnGlnTrpMetGluArgPh 361
1330 1330
361 eGlyProAspThrGlnHisLeuVal.....LeuAsnG 372
||||| ||||| ||||| |||||
1329ACTCAGCATCTGGTCATGCGCGTCACTGGGACTCCAGC 1290
372 luAsnCys.....AlaSerValHisAsnLeuArg 381
::: ||| ||||| |||||
1289 AGGCGCTGGTCCGAGCCACTTGCCTCCCGCTCTGTGCCAGCTCTCCAG 1240
382 ...SerHisLysIleGlnThrGlnLeuAsn...LeuIleHisProAspII 396
::: ||| ::::: ||| ::::: |||
1239 CTCGGTCACGCGGTCTGGAGGAGTCCAGGCTCTCTTCATGCTGCTG 1190
396 ePhe...ProLeuLeuThrSerPheArgCysLysLysGluGlyProThrL 412
||| ||||| ||||| |||||
1189 CTTCTCTCCAGTTCG.....CCGGAAGGCCCGATC 1158
412 euSerValProMetValGlnGlyGluCysLeuLeuLysTyrGlnLeuArg 428
|| ||| ::::: ||| ||||| |||||
1157 TCCTGTCAAGCTCTTCAGTTCTCTGCTGGCTGCTATCACTCGCTG 1108
429 ProArgArgGluTrp.....GlnArgAspAlaIleIleThrCy 441
1107 TAAATGAAGGATTTTGGATTCAATTATTTCGAAGCATTTCTTTGTCTCT 1058
441 sAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGlnG 458
||| ||| ::::: ||| ::::: |||
1057 CAAT...TTCAAATTTCAAAGCTACATATCTGATGTGTGCAACATTTGPAVG 1011
458 InSerValGlnGluIleTyArgArgSerAlaGlnAspGlyProAlaProAla 474
||| ||| ||||| |||||
1010 CTCGTGTTTTTTCTACACTTTCATTCTGCAACAAGAACCTTCITTC 961
475 GluLysArgSer.....GlnTyProGluIleIlePheLeuGlyLys 488
||| ||| ||| ::::: |||
960 GAGCGAGTTGCTCCACTCTTCACAGGTTTACGTGTCGACGCGCGAGC 911
488 rGly.....SerAlaIleProMetLysIleArgAsnValSerA 501
||| ||| ::::: ||| ::::: |||
910 TGGCTCTCTGGGCTTGATCTGCTGGTTGTCCTCTGAAACCGCAGCCA 861
501 laThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAspCysGly 517
::: ||| ::::: |||
860 TAGCGCTTAAACACACAGTCTCGG.....GGC 832
518 GluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspGlnValAspAr 534
||||| ||||| ||||| ::::: |||
831 ATTGACACACTCTGACAAGTGTGCACCTCAACTCGCTCTCAG..... 790

```

534 gValLeuGlyThrLeuAlaAlaValAlaPheValSerHisLeuHisAlaAspH 555
      :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
789 .....GAGAGCTGCAGCTGCATCTGTGAGGCAGACAC.....C 753
551 ishIshThrGlyLeuProSer.....lleLeuLeuGlnArgGluArg... 564
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
752 ACCACGACGAGGACAGTCGGTGTCCTTCGTGTTCTTCGACGCGATCATCGG 703
565 .....AlaLeuAlaSerLeuGlyLysPro..... 572
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
702 AACCTGACTCTTCGACGGCTGCATGTCGCTCCGGTATTTACAGCGCT 653
573 .IeuHisProLeuValValAlaProAsnGlnLeuLysAlaTrpLeuG 589
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
652 TCTCCACAGTCGTCGAGGCTCTTCCTCAAGACCTTTT..... 614
589 lncIlnYrHisAsnGlnCysGlnGluValLeuHisHisIleSerMetIle 605
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
613 .....CTTTCAGTCAGGACGCACAC..... 593
606 ProAlaLysCysLeuGlnGluGlyAlaGlu...lleSerSerProAlaVa 621
      :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
592 ..ATGGAAGTCTCTCAAAATGGCAATCATTTTAAATGACACACGAGA 546
621 lGluArgLeuIleSerSerLeuLeuArgThrCys..... 632
      :||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
545 TGTCCACAGCATTAACCTCTCTGCACAACTCTGCTTTCATTCGACAAATA 496
633 .....AspLeuGluGluPheGlnThrCys..... 640
495 GATCTGAGAGACGAGAATTTCTCTCTTGCAGCAATATCCTTAAACACCT 446
641 ...LeuValArg..... 643
445 TATCTTTAAGATGCTCTCTTGACACGCTGTACATTTTGGACTTGAAGAG 396
644 .....HisCysLysH 647
395 CTCAGACGGCCCGCATGCAGCTCTCGACAGCGGTGCCACACACTCGGT 346
647 isAlaPheGlyCys.....AlaLeuValHisThrSerGlyTrp 659
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
345 CTGCTTCGGCTGCACAGACACAGGTGGCACTTCTCACACTGTACTGT 296
660 LysValValTrpSerGlyAspThrMetProCysGluAlaLeuValArgMe 676
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
295 CCT...CCAGGGTCTCACAAACTTTTCTTGTAAACCTCTTGTTCAGG 249
676 tGlyLysAspAlaThr 681
248 ACAAAACACTGGCGTCC 233
seq_name: /cgml_7/ptodata/1/lna/5B_COMB.seq:US-08-804-227C-7
seq_documentation_block:
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:

```



```
314 luSerPheIleGlnPro.....IleCysGluAsnAlaThrPheGln 327
      ::::: |||||
38085 CGGACTACACCGCGGCCCTCTTCGGGTGGAGACCGCGCTCTTCGG 38134
      ::::: |||||
328 ArgTyrGln..... 330
      ::::: |||||
38135 CTCCTGGAGAGTGGGGCTGTGTCGGGACGCTCCTGGTGGGCGACATCGAT 38184
      ::::: |||||
331 ....GlyLysAlaAspAlaProValAlaLeuValHisMetAlaProA 346
      ::::: |||||
38185 CGGCGTCTGGTGGCGGCTCAGTGGGGGCGTCTCTCTGGCGCGGACG 38234
      ::::: |||||
346 laSerValLeuValAspSerArgTyrGln..... 355
      ::::: |||||
38235 CGGCGCGGTGTCTCCGACGCGCGGCTCATGCGGGCCCTGCGCGGAG 38284
      ::::: |||||
356 .....GlnTrpMetGluArg..... 360
      ::::: |||||
38285 GCGCGGCGATGCGGCGCTGTCAGGCCACCGAGCGGAGCGCGCGCT 38334
      ::::: |||||
361 .....PheGlyProA 364
      ::::: |||||
38335 GGAGCCGTCGCGCGCGGCGGCGGTGTGTCGCGCGGTCAACGCCCG. 38383
      ::::: |||||
364 spThrGlnHisLeuValLeuAsnGluAsnCysAlaSerVal..... 377
      ::::: |||||
38384 .....CAGGCCCTCGTCTCTCGGGAGGAGCGCGCTACTGCGCGG 38428
      ::::: |||||
378 .....HisAsnLeuArgSerHisLysIleGlnThrCl 388
      ::::: |||||
38429 GCGGTGAACTGCGCGCGCGCGGCGGCGGACCAACCGCTGAGGGTGAG 38478
      ::::: |||||
388 nLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgC 405
      ::::: |||||
38479 CCACGCCCTCCACTCACCCGATGAGCGCCCATGCTCGCCGACTTCGCG 38528
      ::::: |||||
405 ysLysLysGlu.....GlyProThrLeuSer...ValPro 415
      ::::: |||||
38529 CGGTGGCGGACACGCTCGACTACCGCGCGCGCGCGCTGCGGTCTCTCC 38578
      ::::: |||||
416 MetValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgArgCl 432
      ::::: |||||
38579 GAAGTACCGCGGCGCTCGCGCGCGCGCGCGCTGAGCGACCGCGGTA 38628
      ::::: |||||
432 uTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleValG 449
      ::::: |||||
38629 CTGGACCGC.....C 38639
      ::::: |||||
449 luAlaLeuGlnLeuProAsnPheGlnSerValGlnGluTyrArgArg 465
      ::::: |||||
38640 AGGTGGCGGACGCGGTGGCGCTTCGCGCGCGCGCGCTGCGC.....ACCGCG 38683
      ::::: |||||
466 SerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrProGl 482
      ::::: |||||
38684 ACCGCCCGGAC..... 38695
      ::::: |||||
482 uIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgAsnV 499
      ::::: |||||
38695 ..... 38695
      ::::: |||||
499 alSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeuAsp 515
      ::::: |||||
38696 ..GCCGCGACCTTCATCGACTCGGCGCGCGCGCTGCTCG..... 38734
      ::::: |||||
516 CysGly.....GluGlyThrPheGl 522
      ::::: |||||
38735 TCGCGCATGGCGGAGAGTCCCTGCGCGGAGCGCGCGCTGCTGCG 38784
      ::::: |||||
522 yGlnLeuCysArgHisTyrGlyAspGlnValAspArgValLeuGlyThrL 539
      ::::: |||||
38785 CCGCGCACTGCGCGCGCGCGCGGCGGCGGACACCGTGTCTCCGGCGG 38834
      ::::: |||||
```

```
539 euAlaAlaValPheValSerHisLeuHisAlaAsp..... 550
      ::::: |||||
38835 CGCGAGCGGTACGTCGCGCGCGGCGCTCGACTGGCGCGCGCTCTAC 38884
      ::::: |||||
551 .....HisHisThrGlyLeuProSerIleLeuLeuGlnAr 562
      ::::: |||||
38885 GCGGCGACGGGACCGCGCGCGCGCTGCGCGCGCTGCGCGCGCTTCCAGCA 38934
      ::::: |||||
562 gGluArgAla...LeuAlaSerLeuGlyLysProLeuHisProLeuLeuV 578
      ::::: |||||
38935 CAGCGCTACTGGCTGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCT...CGCA 38981
      ::::: |||||
578 alValAlaProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGln 594
      ::::: |||||
38982 CGCGCGCGCGCTCC..... 38995
      ::::: |||||
595 CysGlnGluValLeuHisHisIleSerMetIleProAlaLysCysLeuGl 611
      ::::: |||||
38996 .....GTCCGATCCGTGCCG...GAAGCCGAGCA 39021
      ::::: |||||
611 nGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerSerL 628
      ::::: |||||
39022 GGACGGGCGCTG..... 39034
      ::::: |||||
628 euLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuValArgHis 644
      ::::: |||||
39034 ..... 39034
      ::::: |||||
645 CysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrpLysVa 661
      ::::: |||||
39035 .....TGGGCCG 39042
      ::::: |||||
661 lValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMetGlyL 678
      ::::: |||||
39043 CGTGACGCGCGGTGAGTC.....GCCTGCGCGCGCGCGCGCTGCGCG 39086
      ::::: |||||
678 ysAsp.....AlaThrLeuLeuIleHisGlu 686
      ::::: |||||
39087 CGGACGACGCGCGGTATCGAACAACGAACTGCGCGCGGTCTCGCGCACCTG 39136
      ::::: |||||
687 AlaThrLeuAspGlyLeuGluGluAlaValGluLysThr..... 701
      ::::: |||||
39137 GCGCGCTGGCAGC.....CGCAGCGCGCGCGCGCGCGCGCGCG 39180
      ::::: |||||
702 .....HisSerThrThrSerGlnAlaIle.....SerValGlyM 713
      ::::: |||||
39181 CTGCACTACCGCTCACCTGCGAGCGGATCGAGCGAGCGCTGTCAGT 39230
      ::::: |||||
713 etArgMetAsnAlaGluPheIleMetLeuAsnHis.PheSerGlnArgTy 729
      ::::: |||||
39231 TCAGCCCTCGGATCGCTGCTGATGTCGAGCATGGGCGAGCACACGAA 39280
      ::::: |||||
729 rAlaLysValPro 733
      ::::: |||||
39281 TCGCGGAGCGCG 39293
      ::::: |||||
```

seq_name: /cgnl_7/ptodata/1/ina/6_COMB.seq:us-09-320-878-21

seq_documentation_block:

; Sequence 21, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 300622002120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28


```

; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 5970
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-21

alignment_scores:
    Quality: 113.00      Length: 899
    Ratio: 0.298         Gaps: 43
    Percent Similarity: 42.158      Percent Identity: 19.800

alignment_block:
US-09-434-382-2 x US-09-320-878-21  ..
Align seg 1/1  to: US-09-320-878-21  from: 1  to: 5970

      4  LeuCysSerLeuLeuArgSerAlaAla.....GlyArgThrMetSerG1 18
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2955  CTCGCCAGTTACTCCGACGCGCGGGAACACACACGCGTCCGACACCTCCGC 3004
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
18  nGlyArgThrIleSerGlnAlaProAlaArgArg.....GluArgProA 33
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3005  CGTCCGCGACCTGCTGACGCGCTCGCGGCGGCGGCGGAACCTCACCTC 3054
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
33  rGlyAspProLeuArgHisLeuArgThrArgGluLysArgGlyProSer 49
      |||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3055  CCCGCGCGACGCGCGCTCGCGCTCGACGACACACCGCTGCACGACT 3104
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50  GlyCysSerGlyGlyPro....AsnThrValTyrLeuGlnValAlaAla 65
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3105  GGCTGCTCTCCGCTCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3154
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
65  aGlySerArgAspSerGlyAlaAlaLeuTyrValPheSerGluPheAsnA 82
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3155  CCTGACCCACGCGACACCGCGTCCCTG.....A 3183
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
82  rGlyLeuPheAsnCys.....GlyGluGly... 90
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3184  GGTTCCTCTCTACTGCGCGCGCGCCACACGACGCGGCGGCGGCGGCTC 3233
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
91  .....ValGlnArgLeuMetGlnGluHisLysLeuLysValAlaArgLe 105
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3234  GTCATGACCGAGCGGCGCTCGACGCGCGCGGATACCGGGGTTCGCGCG 3283
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
105  uAspAsnIlePheLeuThrArgMetHisTrpSerAsn.....ValGlyG 120
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3284  CGAG.....CTGCTCCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCG 3327
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
120  lYleu.....SerGlyMetIleLeuThrLeuLysGlu.ThrGlyLeuPr 134
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3328  GGCTTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3377
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
134  oLysCysValLeuSerGlyProGlnLeuGluLysTyrLeuGluAlaI 151
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3378  GAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGAA 3427
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
151  leLysIlePheSerGly.....ProLeuLysGlyIleGluLeuAlaVal 165
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3428  CCCGCTGCTCTCGGCGGCGTGCACCGACCGGCGGCGGCGGCGGCGGCG 3462

```

```

166  ArgProHisSerAlaProGlu.....TyrGluAspGluTh 177
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3463  AGGCTCCACACACACCTCTCGCGCGCGCGCGCGCGAGCGCGTCCAC 3512
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
177  rMetThrValTyrGlnIleProIleHisSerGluGlnArgArgGlyLysH 194
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3513  GGCACGCGCTTCGCGCGCGCGCTCTCGCTCGCTCGCTCGCTCGCTCG 3562
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
194  iGlnProTrpGlnSerProGluArgProLeuSerArgLeuSerProGlu 210
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3563  GGCACACCTGGAGACCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCT 3612
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
211  ArgSer.....SerAspSerGluSerAsnGluAsnGluPro.HisLeuP 225
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3613  CGCTCTCTGGTGGAGCGCTCACGGGAACGACCGCGCGCGCTCGAGCG 3662
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
225  roHisGlyValSerGlnArgArg.GlyValArgAspSerSerLeuValVa 241
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3663  CACATCCAGGAGCTCGCGCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 3712
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
241  lAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeu 258
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3713  G.....GACAGCGCGCGATCCGCG 3732
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
258  ysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIle 274
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3733  AGGCGCGCTCCAGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3782
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
275  IleAlaAlaVal.....LysAspGlyLysSerIleThrHisGluG1 288
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3783  CGCGCGCGCGAGCGTGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3832
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
288  yArgGluIleLeuAlaGluGluLeuLysCysThrProAspProGlyAlaA 305
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3833  CGCGCGCGCTGCTGCTC.....GCGGCGG 3855
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
305  laPheValValGluCysProAspGluSerPheIleGlnProIleCys 321
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3856  CC.....CCCGACGCTCGCTCGCGCGCTCGCGCGCTCGCTCGCTC 3887
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
322  GluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAl 338
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3888  GAACGGGCGCGCGACGAG..... 3905
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
338  aLeuValValHisMetAlaProAlaSerValLeuValAspSerArgTyrG 355
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3906  .....GCCCTGTTCGACGCGCGCGCTC. 3926
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
355  lnGlnTrpMetGluArgPheGlyProAspThrGlnHisLeuValLeuAsn 371
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3927  .....GACGACGCGCTTCGCGATCCTCGAG 3950
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
372  GluAsnCysAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGl 388
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3951  .....TTGCGCGCTCGGTCGAGACCGCGCGCGCGCGCGCGCGCG 3982
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
388  nLeuAsnLeuIleHisProAspIlePheProLeuLeuThrSerPheArgC 405
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
3983  GCTGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4026
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
405  ysLysLysGluGlyProThrLeuSerValProMetValGlnGlyCys 421
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
4027  TGAACCGCGCATGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4070
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
422  LeuLeuLysTyrGlnLeuArgProArgGluTrpGlnArgAspAlaI1 438
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
4071  CTCCTGAGCGGTGAACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4114
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
438  eIleThrCysAsnProGluGluPheIleValGluAlaLeuGlnLeuProA 455
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
4115  GATCCGCTGC.....CTCGCTGGTACGCGCGCGCGCGCGCGCGCG 4149
      :||  :||  :||  :||  :||  :||  :||  :||  :||  :||  :||
455  snPheGlnGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyPro 471

```

4150 AGCCGCCGACCGCTGTCCCGCTGGCGCCAGCTCCCGACCAACGATGCC 4199
472 AlaProAlaGluLysArgSerGlnTyrProGluIleIlePheLeuGlyTh 488
4200 TTGGAGCTG.....TCGCTACCCCGGATGGCT 4228
488 rGlySerAlaIleProMetLysIleArgAsnValSerAlaThr..... 502
4229 CGCGCGCTGTCCCGCGCTCTCTGGAGTCCCTGCGCGCACGCCGAGC 4278
503LeuValAsnIleSerProAspThrSerLeuLeu 513
4279 CGAGGGGGTCCGTCCTCCCGTACCGCTCCCGCGGACGCC..... 4322
514 LeuAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAs 530
4322 4322
530 pGlnValAspArgValLeuGlyThrLeuAlaValPheValSerHisL 547
4323GCGCTCCAGCCCGACGCCGCGCTCTTC..... 4349
547 euHisAlaAspHisThrGlyLeuProSerIleLeuLeuGlnArgGlu 563
4350CAGCGGGGC 4358
564 ArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuValValAl 580
4359 CCGGACACGCCCTCGTCCG..... 4379
580 aProAsnGlnLeuLysAlaTrpLeuGlnTyrHisAsnGlnCysGlnG 597
4380CAGGCGGAAC 4389
597 luValLeuHisHisIleSerMetIleProAlaLysCysLeuGlnGly 613
4390 AGATCTG.....CAGGGC 4403
614 AlaGluIleSerSerProAlaValGluArgLeuIleSerSerLeuAr 630
4404 TCCCGGCTGTCCGAGGAGACGTACGAGGCCCTGGAGACGCCCTCTTGT 4453
630 gThrCysAspLeuGluGluPheGlnThrCysLeuValArgHisCysLysH 647
4454 CCGTCCACGCCGACCGGCTCGCGGCGGTTC..... 4490
647 isAlaPheGlyCysAlaLeuVal.....HisThrSer 657
4491TGGTCGACGCCCTCTCGCGAGCGCTGGAGCGCGGTCTC 4535
658 GlyTrpLysValItyrSerGlyAspThrMetProCysGluAlaLeuVa 674
4536 GCCTGGAGCGGCTCTCGCGCGACCCGG..... 4565
674 lArgMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluA 691
4566GCGATCGATCGCTCGCGGACCTCCCGA 4599
691 spGlyLeuGluGluAlaValGluLysThrHisSerThr...ThrSer 706
4600 CGCGCGGAGCGGCGGAGCTGGCGCTCTCCACGCGCGCGGAGAGC 4649
707 GlnAlaIleSerValItyrMetArgMetAsnAlaGluPheIleMetLeuAs 723
4650 TGGGGCTCGCGTGGCATGCCCTCTCCGGCTGCTGCTCGCTCGCTGCAC 4699
723 nHisPhe.....SerGlnArgTyrAlaLysValProLeuP 735
4700 GGAGCGCGCGAGTACGAACAGCGGCGGTCTCGCGGACGCGGTGC 4749
735 heSerProAsnPheSerGluLysValGlyValAlaPheAspHis..... 749

4750 CGGACCGGATGTTCGACTCGCGGACGGCATGGAGTACATGCACGCCCGG 4799
750MetLysValCysPheGlyAspPh 757
4800 GGCGCGCTACTGGCTGGCGANCGCGGCTGCACGCGCGCTGGCGGAGTT 4849
757 eProThrMetProLysLeu.....I 764
4850 CATGTCTCTCGGGGAGATCTGGGCAGCTTGAACCTCGACACGCCCTCGA 4899
764 leProProLeuLysAlaLeuPheAlaGlyAspIleGluMetGluGlu 780
4900 TCGTGCCTGGCGGACCTCCGCGCGGAGGTGTACCTGGGCTCGGCAAC 4949
781 ArgArgGluLysArgGluLeuArgGlnValArgAlaAlaLeuLeuSer... 796
4950 CGCCAGAGCCCGGCGCTGGCCGAGGCCACCTCGCCCTGTGTGGCGCC 4999
797ArgGluLeuAlaGlyGlyL 803
5000 CGGGCGCTCCCGCACCCGCGGTCTCACCTGGCGGTCTTGGCGCGCGG 5049
803 euGluAspGlyGluProGlnGlnLysArgAlaHisThrGlu 816
5050 TG...GACGGCCAG...CAGCGGAGCGGTGCACGCCGAG 5084

OM of: US-09-434-382-2 to: EST:* out_format : pfs

Date: Feb 18, 2001 4:09 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODE=frame+P2n.model -DEV=xlp
-O=/cgn1_1/USPTO_spool/US09434382/runat_16022001_105748_29671/app_query.fasta_1.895
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-ALIGN=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09434382.cgn1_1_4002 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-434-382-2

Query length: 826

Database: EST:*

Database sequences: 7991742

Database length: -79123438

Search time (sec): 1395.210000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_est75:BE795820	+ 1242.50	2207.27	6.2e-114	761	! BE795820 601590856F1 NIH_MGC_7
gb_est69:BE260495	+ 1180.00	2096.49	9.2e-108	676	! BE260495 601150702F1 NIH_MGC_19
gb_est75:BE744197	+ 1165.00	2065.13	5.1e-106	949	! BE744197 601577168F1 NIH_MGC_9
gb_est73:BE619259	+ 1164.50	2066.08	4.5e-106	823	! BE619259 601473130F1 NIH_MGC_68
gb_est75:BE744876	+ 1158.50	2052.98	2.4e-105	984	! BE744876 601576324F1 NIH_MGC_9
gb_est76:BE867512	+ 1134.50	2048.28	4.4e-105	812	! BE867512 601443010F1 NIH_MGC_65
gb_est77:BE892893	+ 1131.00	2008.20	7.6e-103	992	! BE747163 601577254F1 NIH_MGC_9
gb_est77:BE900936	+ 1123.00	1992.03	6.0e-102	790	! BE794311 601591442F1 NIH_MGC_7
gb_est77:BE900936	+ 1109.00	1987.39	1.1e-101	745	! BE900936 601674206F1 NIH_MGC_21
gb_est77:BE902696	+ 1086.50	1928.68	2.0e-98	664	! BE902696 601677393F1 NIH_MGC_21
gb_est70:BE383336	+ 1066.00	1891.30	2.5e-96	692	! BE383336 601298249F1 NIH_MGC_19
gb_est69:BE260626	+ 1005.00	1777.73	5.2e-90	938	! BE260626 601146116F1 NIH_MGC_19
gb_est75:BE742908	+ 1000.50	1773.87	8.6e-90	677	! BE742908 601574609F1 NIH_MGC_9
gb_est69:BE304720	+ 996.00	1767.09	2.0e-89	612	! BE304720 601106236F1 NIH_MGC_15
gb_est75:BE743831	+ 977.00	1727.38	3.3e-87	940	! BE743831 601577742F1 NIH_MGC_9
gb_est72:BE537086	+ 975.00	1727.37	3.3e-87	713	! BE537086 601063325F1 NIH_MGC_10
gb_est3:AA184645	+ 935.00	1657.98	2.4e-83	588	! AA184645 mt58d10.r1 Soares.thym
gb_est4:AA243700	+ 897.00	1591.01	1.3e-79	531	! AA243700 zr68q08.sl Soares.NbHM
gb_est68:BE250309	+ 897.00	1590.00	1.5e-79	574	! BE250309 600943455F1 NIH_MGC_17
gb_est41:AW206103	+ 892.00	1581.91	4.2e-79	536	! AW206103 UI-H-BL1-afx-f12-0-UT
gb_est77:BE914535	+ 882.00	1562.11	5.3e-78	617	! BE914535 601665633F1 NCI_CGAP_1
gb_est49:AW711657	+ 871.00	1544.79	4.9e-77	511	! AW711657 hn59h06.x1 NCI_CGAP_1
gb_est46:AW575677	+ 857.00	1520.09	1.2e-75	493	! AW575677 UI-HF-BM0-adi-b-07-0-UT
gb_est70:BE409312	+ 855.50	1513.02	2.9e-75	691	! BE409312 601300940F1 NIH_MGC_21
gb_est37:AV615441	+ 841.00	1490.08	5.5e-74	543	! AV615441 AV615441 Bos taurus ad
gb_est70:BE371842	+ 828.00	1461.65	2.1e-72	803	! BE371842 601217643F1 NCI_CGAP_1
gb_est73:BE615669	+ 825.00	1455.19	4.8e-72	872	! BE615669 601279347F1 NIH_MGC_39
gb_est41:AW175923	+ 818.00	1449.51	1.0e-71	512	! AW175923 RCI-BT0077-020999-002
gb_est7:AA455121	+ 809.50	1434.72	6.7e-71	493	! AA455121 zx78c04.sl Soares ovar
gb_est5:AA311855	+ 809.00	1433.62	7.7e-71	501	! AA311855 EST182568 Jurkat T-cell
gb_est4:AA291670	+ 807.00	1430.74	1.1e-70	474	! AA291670 zt37d04.sl Soares ovar
gb_est17:AI200296	+ 798.00	1397.24	8.1e-69	451	! AI200296 qf86b12.x1 Soares_feta
gb_est10:AA531139	+ 771.50	1364.37	5.5e-67	578	! AA531139 v105g04.y1 Barstead md
gb_est8:AA634909	+ 762.00	1350.66	3.2e-66	446	! AA634909 ab27h02.r1 Stratagene
gb_est22:AI553139	+ 760.00	1345.37	6.3e-66	508	! AI553139 v105g04.y1 Barstead md
gb_est5:AA310236	+ 754.00	1336.72	1.9e-65	431	! AA310236 EST181085 Jurkat T-cell
gb_est10:AA679618	+ 741.00	1312.74	4.1e-64	452	! AA679618 ag72c12.sl Gessler WIL
gb_est7:AA409793	+ 740.00	1311.44	4.9e-64	435	! AA409793 EST015133 Mouse 7.5 opd
gb_est72:BE540045	+ 735.50	1295.20	3.9e-63	816	! BE540045 601060822F2 NIH_MGC_10

gb_est86:W27286 - 735.00 1293.70 4.8e-63 855 ! W27286 28h1 Human retina cDN
gb_est20:AI468143 + 734.00 1301.08 1.8e-63 421 ! AI468143 tf92g05.x1 NCI_CGAP
gb_est84:R87541 + 730.50 1292.11 5.8e-63 518 ! R87541 ym89b04.r1 Soares adu
gb_est42:AW304130 + 702.00 1244.11 2.8e-60 404 ! AW304130 xs13e05.x1 NCI_CGAP

seq_name: gb_est75:BE795820

seq_documentation_block:
LOCUS BE795820 761 bp mRNA EST 20-SEP-2000
DEFINITION 601590856F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945085 5',
mRNA sequence.
ACCESSION BE795820
VERSION BE795820.1 GI:10217018
FEATURES
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 761)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTF
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM802 row: p column: 14
High quality sequence stop: 761.

FEATURES
source
1..761
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3945085"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 172 a 206 c 219 g 164 t
ORIGIN

alignment_scores:
Quality: 1242.50 Length: 253
Ratio: 5.010 Gaps: 1
Percent Similarity: 98.024 Percent Identity: 96.838

alignment_block:
US-09-434-382-2 x BE795820 ..

Align seg 1/1 to: BE795820 from: 1 to: 761

494 MetLysIleArgAsnValSerAlaThrLeuValAsnIleSerProAsp510
|||||

2 ATGAAGATTCGAATGTGTCAGTCCACACTTGTCAACATAGCCCGACAC 51
|||||

510 rSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysArgH 527
|||||

52 GTCTCTGCTACTGACTGTGTGTGAGGGGACCGTTTGGCAGCTGTGCCGTC 101
|||||

527 iStyRGlyAspGlnValAspArgValLeuGlyThrLeuAlaAlaValPhe 543
|||||

```
|||||
102 ATTACGAGACACAGGTGGACAGGCTCTGGGACACCTGGCTGCTGTTT 151
544 ValSerHisLeuHisAlaAspHisHisThrGlyLeuProSerIleLeuLe 560
|||||
152 GTGTCCTCCACCTGCACGCAGATCACACACGGGCTTCCCAAGTATCTTGC 201
560 uGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProLeuL 577
|||||
202 GCAGAGAGAACGGCGCTTGGCATCTTTGGGAAGCGGCTTCAACCTTTGC 251
577 euValValAlaProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsn 593
|||||
252 TGGTGGTTGCCCTCCCAACACAGCTCAAGCCCTGGCTCCAGCAGTACCA 301
594 GlnCysGlnGlnValLeuHisHisIleSerMetIleProAlaLysCyste 610
|||||
302 CAGTGGCAGGAGTCTCGCACCATCATGATGATTCCTGCCAAATGCGCT 351
610 uGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleSerS 627
|||||
352 TCAGGAAGGGGTGAGATCTCCAGTCTCGAGTGGAAAGATTGATCAGTT 401
627 erLeuLeuArgThrCysAspLeuGluPheGlnThrCysLeuValArg 643
|||||
402 CCGTGTGGCAACATGTGATTG.GAAGAGTTTCAGACCTGTCTGGTGGCG 450
644 HisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTrp 660
|||||
451 CACTGCAAGCATGCGTTTGGCTGTGGCTGGTGCACACCTCTGGCTGGAA 500
660 sValValTyrSerGlyAspThrMetProCysGluAlaLeuValArgMetG 677
|||||
501 AGTGTGCTATTCCGGGGACACCATGCGCTCGAGGCTCTGGTCCGGATGG 550
677 lYlYAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGlyLeu 693
|||||
551 GGAAGATGCCACCTCTCTGATACATGAAGCCACCTTGAAGATGGTTG 600
694 GluGluGluAlaValGluLysThrHisSerThrThrSerGlnAlaIleSe 710
|||||
601 GAAGAGGAAGCAGTGGAAAGACACACAGCACAACGTCCTCCCAAGCCAT 650
710 rValGlyMetArgMetAsnAlaGluPheIleMetLeuAsnHisPheSerG 727
|||||
651 CGTGGGATGCGGATGAAGCGGAGTTCATTATGCTGAACCACTTCAGC. 699
727 InArgTyrAlaLysValProLeuPheSerProAsnPheSerGluLysVal 743
|||||
700 ..AGCTATGCCAAGTCCCTCTTCAGCCCACTTCAGCCGAGAAAGTG 747
744 GlyValAla 746
|||||
748 GGAGTTGCT 756
```

seq_name: gb_est69:BE260495

```
seq_documentation_block:
LOCUS BE260495 676 bp mRNA EST 13-JUL-2000
DEFINITION 601150702F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3503184 5',
mRNA sequence.
ACCESSION BE260495
VERSION BE260495.1 GI:9131807
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 676)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
```

Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CMI76 row: d column: 01
High quality sequence stop: 672.

FEATURES

Location/Qualifiers

source

1..676

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3503184"

/clone_lib="NIH_MGC_19"

/tissue_type="neuroblastoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: brain; Vector: pOT87; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGCAGGAG(G). Library constructed by Ling Hong

in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT 154 a 207 c 176 g 139 t

ORIGIN

alignment_scores:

Quality: 1180.00 Length: 225

Ratio: 5.244 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-434-382-2 x BE260495 ..

Align seg 1/1 to: BE260495 from: 1 to: 676

```
365 ThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuAr 381
|||||
1 ACCCAGCATTGTGCTGTAATGAGAAGTGTGCTCAGTTTCACAACTTCG 50
381 gSerHisLysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePheP 398
|||||
51 CAGCCACAGATTCAACCCAGCTCAACCTCATCCACCCGGACATCTTC 100
398 roLeuLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerVal 414
|||||
101 CCTGCTCACCAGTTTCGCTGTAAGAGAGGGGGCCCACTTCAGTG 150
415 ProMetValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgAr 431
|||||
151 CCCATGGTTTCAGGGTCAATGCTCTCAAGTACCAGCTCCCTCCAGGAG 200
431 gGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheIleV 448
|||||
201 GGAGTGGCAGAGGGATGCCATTATTACITGCAATCTCGAGGAATTCAT 250
448 aLGlAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrArg 464
|||||
251 TTGAGGGCGTGCAGCTTCCCAACTTCCAGCAGAGCGTGCAGGAGTACAG 300
465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrPr 481
|||||
301 AGAGTGGCAGAGCGGCCACCCAGCAGAGAGAAAGATCAGTACCC 350
481 oGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498
|||||
351 AGAAATCATCTCTCTTGGAAACAGGTCTGCCATCCGATGAAGATTCGAA 400
498 snValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeu 514
```

```

|||||
401 ATGTCAGTCCACACTGTGCAACATAGCCCGACACGCTCTCGTACTG 450
515 AspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspG1 531
|||||
451 GACTGTGGTGAGGCGACATTTGGGACGCTGTGCGTCATTACGAGACCA 500
531 nValAspArgValLeuGlyThrLeuAlaValPheValSerHisLeuH 548
|||||
501 GTGGACAGGGTCTGGGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 550
548 isAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGlu 564
|||||
551 ACCGATCACCACAGCGGCTTCCCAAGATATCTTGCTGCAGAGAACGC 600
565 AlaLeuAlaSerLeuGlyLysProLeuHisProLeuValValAlaPr 581
|||||
601 GCCTTGGCATCTTTGGGAAGCGCTTACCCTTTGCTGGTGTGCCCC 650
581 oAsnGlnLeuLysAlaTrpLeuGln 589
|||||
651 CAACCAAGCTCAAGCGCTGGCTCCAG 675

```

seq_name: gb_est75:BE744197

```

seq_documentation_block:
LOCUS BE744197 949 bp mRNA EST 15-SEP-2000
DEFINITION 601577168F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838121 5',
mRNA sequence.
ACCESSION BE744197
VERSION BE744197.1 GI:10158189
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 949)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue procurement: DCTD/bnp
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM524 row: g column: 18
High quality sequence stop: 668.

```

```

FEATURES
    source
        1..949
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:3838121"
            /clone_lib="NIH_MGC_9"
            /tissue_type="adenocarcinoma cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: ovary; Vector: pORF7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5',
            adaptor: GGCACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 225 a 263 c 281 g 179 t 1 others

alignment_scores:
Quality: 1165.00 Length: 278

```

Ratio: 4.447 Gaps: 6
Percent Similarity: 94.245 Percent Identity: 90.647
alignment_block:
US:09-434-382-2 x BE744197
Align seg 1/1 to: BE744197 from: 1 to: 949
225 ProHisGlyValSerGlnArgArgGlyValArgAspSerSerLeuValVa 241
|||||
2 CCACATGGTGTAGCCAGAGAAGGGGTGAGGACTCTTCCCTGGTGGT 51
|||||
241 lAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeu 258
|||||
52 AGCTTTTCATCTGTAAGCTTCACTTAAAGAGAGGAACTTCTTGGTGTCA 101
|||||
258 ySalAlaGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIle 274
|||||
102 AAGCAAGAGGATGGGCTCCAGTTGGGACAGTCCCATCGCTCCCATC 151
|||||
275 lIleAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGlu 291
|||||
152 ATTGCTGCTGCTCAAGGACGGGAAAGCATCACTCATGAAGAAGAGAGAT 200
|||||
291 eLeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheVal 308
|||||
201 TTGGCTGAAGAGCTGTGTACTCTCTCCAGATCTCTGGTGTGTC. 249
|||||
308 aIValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAla 324
|||||
250 TGGTAGAATGTCCAGATGAAGCTTCAATCAACCCATCTGTGAGAATGCC 299
|||||
325 ThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValVa 341
|||||
300 ACCTTTTCAGAGGTACCAAGGAAGGACAGATGCCCGCTGGCTGGTGGT 349
|||||
341 lHisMetAlaProAlaSerValLeuValAspSerArgTyrGlnGlnTrpm 358
|||||
350 TCACATGGCCCCAGCATCTGTGCTGTGGACAGCAGGTACCAGCAGTGA 399
|||||
358 etGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCys 374
|||||
400 TGGAGAGGTTTGGGCGCTGACACCCAGCAGTGTGCTCAATGAGAAGTGT 449
|||||
375 AlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLe 391
|||||
450 GCCTCAGTTCCACACCTTCGACGCCACAGATTCAACCCAGCTCAACCT 499
|||||
391 uIleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLys 408
|||||
500 CATCCACCGGACATCTTCCCTGCTCACAGTTTCCGCTGTAAGAAGG 549
|||||
408 luGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeuLys 424
|||||
550 AGGGCCCCACCTCAGTGTGCCCATGTTGAGGGTGAATGCTCTCTCAAG 599
|||||
425 TyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaIleIleThr.C 441
|||||
600 TACCAGTCTCGTCCAGG.AGGAGTGGCAGAGGATGCCATTATTACTTT 648
|||||
441 yAsnProGluGluPheIleVal...GluAlaLeuGln...LeuProAsn 455
|||||
649 GCAATCTCTGAGGAATTCATAGCTTGAGGGCGGCTGCAAGGGCTTCCCAAC 698
|||||
456 PheGlnGln.SerValGlnGluTyrArgArgSerAla...GlnAspGlyp 471
|||||
699 CTCACGAGGAGGCTGCGAGGAGTACCCCGAGGGAGTCCCGCAGGACGCC 748
|||||
471 roAlaProAla...GluLysArgSerGlnTyrProGluIleIlePheLeu 486
|||||
749 CAGCCCCAGTCAGAAAGAAAGAGTACCGCACCAGCAAAATCATTTCTCCT 798
|||||
487 GlyThrGlySerAlaIleProMetLysile 496

```

```

799 GGAACAGTCTGGACCCCGGATGAGATC 828
seq_name: gb_est73:BB619259
seq_documentation_block:
LOCUS BB619259 823 bp mRNA EST 24-AUG-2000
DEFINITION 601473130F1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3876223 5',
mRNA sequence.
ACCESSION BB619259
VERSION BB619259.1 GI:9890197
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 823)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC623 row: k column: 08
High quality sequence stop: 695.
FEATURES
source location/Qualifiers
1..823
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3876223"
/clone_lib="NIH_MGC_68"
/tissue_type="large cell carcinoma"
/lab_host="DHI0B (phage-resistant)"
/notes="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 185 a 219 c 259 g 160 t
ORIGIN

alignment_scores:
Quality: 1164.50 Length: 284
Ratio: 4.603 Caps: 3
Percent Similarity: 89.085 Percent Identity: 84.859

alignment_block:
US-09-434-382-2 x BB619259 ..
Align seg. 1/1 to: BB619259 from: 1 to: 823

519 GlyThrPheGlyGlnLeuGlnCysArgHisTyrGlyAspGlnValAspArgVa 535
|||||
3 GGCACAGTTGGCAGCTGTGCGTCATTACGAGACACAGGTGGACAGGT 52
535 lLeuGlyThrLeuAlaAlaValPheValSerHisLeuHisAlaAspHisH 552
|||||
53 CTGGGACCTTGGCTACTGTGTT.GTGTCACCTGCACGACGATCACC 101
552 lThrGlyLeuProSerIleLeuLeuGlnArgGluArgAlaLeuAlaSer 568
|||||
102 ACACGGGCTTCCAGATATCTTGCTGCAGAGACAGCGCTTGCGATCT 151
569 LeuGlyLysProLeuHisProLeuLeuValValAlaProAsnGlnLeuLy 585
|||||
152 TTGGGAAGCGCGCTTCAACCTTTGCTGTGTGTCGCCCAACACGACTCAA 201
```

```

585 sAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnGlnValLeuHisHisI 602
|||||
202 AGCTTGGCTCCAGCAGTACCAACACAGTGCAGGAGGTCTGCACCACA 251
602 lSerMetIleProAlaLysCysLeuGlnGlnGluGlyAlaGluLeuSer 618
|||||
252 TCAGTATGATTCCTGCCAATCCTTCAGAAAGGGGCTGAGATCTCCAGT 301
619 ProAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGl 635
|||||
302 CCTGCAGTGGAAAGATTGATCAGTTCGCTGTCGGAACATGATGATTGA 351
635 uGluPheGlnThrCysLeuValArgHisCysLysHisAlaPheGlyCysA 652
|||||
352 ACAGTTTTCAGACCTCTCTGCTGGCGCACATGCAAGCATGCGTTTGGCTG 401
652 lAlaLeuValHisThrSerGlyTyrLysValValTyrSerGlyAspThrMet 668
|||||
402 CGCTGGTGCACACCTCTGCTGGCTGGAAGTGTCTATTCCGGGGACACCAT 451
669 ProCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHi 685
|||||
452 CCCTCGGAGGCTCTGCTCGGATGGGAAAGATGCCACCTCCTGATACA 501
685 sGluAlaThrLeuGluAspGlyLeuGluGluAlaValGluLysThrH 702
|||||
502 TGAAGCCACCTTGAAGATGTTTGAAGAGGCAAGCAGG.GAAAAGACAC 550
702 lSerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGlu 718
|||||
551 ACAGCACCAACGTCCCAAGCCATCAGCGTGGGATCGGATGAACGCGGAG 600
719 PheIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProLeuPh 735
|||||
601 TTCATTATGCTGAACCATCTTCAGCCAGCGCTATGCCAAGGTCCCTCTT 650
735 eSerProAsnPheSerGluLysValGlyValAlaPheAspHisMetLysV 752
|||||
651 CAGCCCAACCTTCAGCGAGAAAGTGGGAGTGGCTTGACCATCAT...GAAG 697
752 aLysPheGlyAspPheProThrMetProLysLeuIleProProLeuLys 768
|||||
698 GTCTGCTGGGACATTTCCAAACAATGCCAAGGTGATTCCCAATGAAG 747
769 AlaLeuPheAlaGlyAspIleGluMetGluGluArgArgGluLysAr 785
|||||
748 CTGTG...TGGGGGAC..... 762
785 gGluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAlaGlyG 802
|||||
763 .....ACGAGGCAAGGGGGCGGGCGGAAACGGAGCTGGCAAGG 805
802 ly 802
||
806 GG 807

seq_name: gb_est75:BE744876
seq_documentation_block:
LOCUS BE744876 984 bp mRNA EST 15-SEP-2000
DEFINITION 601576324F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3837533 5',
mRNA sequence.
ACCESSION BE744876
VERSION BE744876.1 GI:10158868
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 984)
AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
```

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/BTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM522 row: o column: 06
High quality sequence stop: 726.

FEATURES

Location/Qualifiers
source
1. 984

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3837533"
/clone_lib="NIH_MGC_9"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGCAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 227 a 259 c 316 g 182 t
ORIGIN

alignment_scores:

Quality: 1158.50 Length: 253
Ratio: 4.634 Gaps: 2
Percent Similarity: 98.814 Percent Identity: 96.838

alignment_block:

US-09-434-382-2 x BE744876 ..

Align seg 1/1 to: BE744876 from: 1 to: 984

225 ProHisGly.ValSerGlnArgGlyValArgAspSerLeuValV 241
|||||
2 CCACATGGTGGTAGCAGAGAAGGGGCTCAGGACTCTCCCTGGTGC 51
241 aAlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuVal 257
|||||
52 TAGCTTTCATCTGTAGCTTCACTTAAGAGAGGAACTCTTGGTGCTC 101
258 LysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaPro 274
102 AAAGCAAGAGATGGGCTCCAGTTGGACAGCTGCCATGCCAT 151
274 eileAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGlu 291
152 CATGTGCTGTCTCAGGAGCGGAAAGCATCACTCATGAAGGAAGAGAGA 200
291 leuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheVal 307
201 T.TTGCTGAAGAGCTGTACTCTCCAGATCCTGGTGCTGC.TTTGTG 249
308 ValValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAl 324
250 GTGGTAGATGTCACAGATCAAGCTTTCATCAACCATCTGTGAGATGC 299
324 aThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValV 341
300 CACTTTCAGAGGTACCAAGGAAGGAGATGCCCGCTGGCTGCTG 349
341 aHisMetAlaProAlaSerValLeuValAspSerArgTyrGlnGlnTrp 357
350 TTCATATGGCCCGAGCATCTGTGCTTGTGGACAGCAGGTACCAGCAGTGG 399

358 MetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCy 374
|||||
400 ATGCAGAGGTT.GGGCTGTACACCCAGCACTTGTCTGAATGAGAACTG 448
374 sAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnL 391
|||||
449 TGCTCAGTTCAACAGCTTGGCAGCCACAGATTCAAAACCCAGCTCAAC 498
391 eulIleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLys 407
|||||
499 TCATCCACCGGACATCTCCCTGCTCACCAGTTTCCGCTGTAAAGAG 548
408 GluGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeu 424
|||||
549 GAGGCGCCCACTCAGTGTGCCAGTGTTCAGGGTGAATGGCTCTCTCAA 598
424 sTyrGlnLeuArgProArgArgGluTrpGlnArgAspAlaIleIleThrC 441
599 GTACCAAGCTCCGCCAGGAGGAGTGGCAGAGGGATGCCATTATTACTT 648
441 ysAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGln 457
|||||
649 GCATCTCTGAGGATTCATAGTTGAGCGCTGAG.CTTCCCAACTTCCAG 697
458 GlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlaProAl 474
|||||
698 CACAGCGTGCAGGAGTACAGGAGGAGTGCAGAGGCC...AGCCAGC 744
474 aGluLys 476
|||||
745 AGAGAAA 751

seq_name: gb_est76:BE867512

seq_documentation_block:

LOCUS BE867512 812 bp mRNA EST. 27-SEP-2000
DEFINITION 601443010F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3847226 5',
mRNA sequence.

ACCESSION BE867512 GI:10316288

VERSION BE867512

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 812)

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM548 row: c column: 03

High quality sequence stop: 686.

Location/Qualifiers

1. 812

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3847226"

/clone_lib="NIH_MGC_65"

/tissue_type="adecarcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;

Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

Average insert size 1.8 kb. Library constructed by Life

BASE COUNT 200 a 216 c 220 g 176 t
 ORIGIN
 alignment_scores:
 Quality: 1154.50 Length: 270
 Ratio: 4.655 Gaps: 5
 Percent Similarity: 91.852 Percent Identity: 90.000
 alignment_block:
 US-09-434-382-2 x BE867512 ..

Align seg 1/1 to: BE867512 from: 1 to: 812

242 AlaPheIleCysLysLeuHisLeuLysArgGlyAsnPheLeuValLeuLys 258
 |||||
 1 GCTTTCATCTGTAAGCTTCACCTAAAGAGAGAAACTCTTGGTGCTCA 50
 258 sAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProIleI 275
 |||||
 51 AGCAAAGAGATGGCCCTCCAGTTGGGACAGCTGCCATCGTCCCATCA 100
 275 leAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGluIle 291
 |||||
 101 TTGCTGCTGTCAGGACCGGAAAGCATCATCATGAAGGAGAGAGATT 150
 292 LeuAlaGluGluLeuCysThrProProAspProGlyAlaAlaPheValVa 308
 |||||
 151 TTGGCTGAAGAGCTGTGTACTCTCCACAGATCTGTGTGTGT 199
 308 lvaIgluCysProAspGluSerPheIleGlnProIleCysGluAsnAla 325
 |||||
 200 GGTAGAATGTCAGATGAAGCTTCATTCAACCCATCTGTGAGAAATGCCA 249
 325 hrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValVal 341
 |||||
 250 CCTTTCAGAGGTACACGAAAGGAGATGCCCGTGGCTTGTGTGT 299
 342 HisMetAlaProAlaSerValLeuValAspSerArgTyrGlnGlnTrpMe 358
 |||||
 300 CACATGGCCCCAGCATCTGTCTGTGGACAGCAGGTACAGCAGTGGAT 349
 358 tGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCysA 375
 |||||
 350 GGAGAGGTTGGGCTGACACCCAGCACTTGTCTGTAATGAGAACTGTG 399
 375 laSerValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLeu 391
 |||||
 400 CCTAGTTTCAACACTTCGACGCCACAAAGATTCAACCCAGCTCAACCTC 449
 392 IleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLysG1 408
 |||||
 450 ATCCACCCGGACATCTTCCCTGCTCACCAGTTTCCGCTGTAAGAAGGA 499
 408 uGlyProThrLeuSerValProMetValGlnGlyCysLeuLeuLysT 425
 |||||
 500 GGGCCCCACCTCAGTGGCCATGGTTGAGGGTGAATGCTCTCAAGT 549
 425 yrGlnLeuArgProArgArg.GluTrp.GlnArgAspAlaIleIleThr. 440
 |||||
 550 ACCAGCTCCGTCACGAGGCGGATGGGCGAGGAGATGCCATTATTACTC 599
 441 CysAsnProGluGluPheIleValIcIuAlaLeuGlnLeuProAsnPheG1 457
 |||||
 600 TGCAATCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCACTTCAG 648
 457 nGlnSerValGlnGluTyrArgArgSerAlaGlnAspGlyProAlaProA 474
 |||||
 649 .CAGAGCGGTGCAGAGTACAGGAGAGTGGCCGACGAGCGGCCAGCCAG 698
 474 laGluLysArgSerGlnTyrProGluIleIlePheLeuGlyThrGlySer 490
 |||||

699 CAGAGAAAAGAGTCTAGTACCAGGAATCATC.....TCCTGGGACGGT 742
 491 AlalleProMetLysIle.....Argas 498
 |||||
 743 CTGCATCCGATCAAGATCGAATGTCGGGCCACTGTTAATATTAGCCGAAA 792
 498 nVal 499
 |||||
 793 CGTT 796

seq_name: gb_est75:BE747163

seq_documentation_block:

LOCUS BE747163 992 bp mRNA EST 15-SEP-2000
 DEFINITION 601577254F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3838237 5',
 mRNA sequence.

ACCESSION BE747163

VERSION BE747163.1 GI:10161155

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 992)

AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCMS24 row: 1 column: 14

High quality sequence stop: 781.

FEATURES

source

1..992

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3838237"

/clone_lib="NIH_MGC_9"

/tissue_type="adenocarcinoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:

ECORI; cDNA made by oligo-dT priming. Directionally

cloned into EcoRI/XhoI sites using the following 5'

adaptor: GGACGAG(G). Size-selected >500bp for average

insert size 1.8kb. Library constructed by Ling Hong in

the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 225 a 287 c 313 g 167 t

ORIGIN

alignment_scores:

Quality: 1140.00 Length: 275

Ratio: 4.488 Gaps: 5

Percent Similarity: 92.364 Percent Identity: 89.818

alignment_block:

US-09-434-382-2 x BE747163 ..

Align seg 1/1 to: BE747163 from: 1 to: 992

331 GlyLysAlaAspAlaProValAlaLeuValHisMetAlaProAlaSe 347

1 GGAAAGCAGATGCCCCGTGGC.CTGGTGGTTACATGGCCCCCAGCATC 49

347 rValLeuValAspSerArgTyrGlnGlnTrpMetGluArgPheGlyProA 364


```

|||||
50  TGTGCTTGGACAGCAGGATACAGCAGTGGATGGAGGTTT.GGCGCTG 98
|||||
364  spThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeu 380
|||||
99  ACACCCAGCAGCTTGGTCTCTGAATGAGAAGTGTGCTCAGTTCAACACCTT 148
|||||
381  ArgSerHisIysIleGlnThrGlnLeuAsnLeuIleHisProAspIlePh 397
|||||
149  CGCAGCCACAAGATTCAAAACCCAGCTCAACCTCATCCACCGGACATCTT 198
|||||
397  eProLeuLeuThrSerPheArgCysIysIysGluGlyProThrLeuSerV 414
|||||
199  CCCCCTGCTCACCAAGTTCGCTGTAGAGAGGAGGCGCCACCCCTCAGTG 248
|||||
414  alProMetValGlnGlnGlyGluCysLeuLeuIysTyrGlnLeuArgProArg 430
|||||
249  TGCCCATGGTTTCAGGCTGAATGCTCTCAAGTACCAGCTCCGTCGCCAGG 298
|||||
431  ArgGluTrpGlnArgAspAlaIleIleThrCysAsnProGluGluPheII 447
|||||
299  AGGAGTGGCAGAGGATGCCATTATTTACTTGCATCTCCTGAGGAATTCAT 348
|||||
447  eValGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrA 464
|||||
349  AGTTGAGGCGCTCCAGCTTCCCAACCTTCCAGCAGAGCGTGCGAGGATACA 398
|||||
464  rArgSerAlaGlnAspGlyProAlaProAlaGlnLysArgSerGlnTyr 480
|||||
399  GGAGGAGTGGCAGGACGG.CCAGCCCCAGCAGCAGAGAAAGAACTCAGTAC 447
|||||
481  ProGluIleIlePheLeuGlyThrGlySerAlaIleProMetIysIleAr 497
|||||
448  CCAGAAATCATCTCTCTTGAACAGGCTCTGCCATCCGATGAAGATTCCG 497
|||||
497  gAsnValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuL 514
|||||
498  AAATGTCAGTGCCACCACTTGTCAACATAGCCCGGACACGCTTTTGTCTAC 547
|||||
514  euAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAsp 530
|||||
548  TGGACTGTGTGAGGCGACATTTGGGCGAGCTGTGCCGTCTATTACGGAGAC 597
|||||
531  GlnValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLe 547
|||||
598  CAGGTGGACAGGCTCTGGGCAACCTGGTGTGTGTGTGTGTGTGTGTGTGT 647
|||||
547  uHisAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluA 564
|||||
648  GCACGCGAGAT...CACACGCGCTGCAAGTATCTTCTGCAGAGAGAAC 694
|||||
564  rAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeuValValAla 580
|||||
695  CGCGCTG....GATCTGGGAAAGCGGTTCACCTTAGTGG...TGGTGC 737
|||||
581  ProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnG 597
|||||
738  CCCAACGAGTCAAAAGGGG.....TCAGCAGTACACAAACGAGTGCAGGAG 781
|||||
597  uValLeuHisHisIleSerMetile 605
|||||
782  GTCTGCC...CACATAGGATGATC 803

```

seq_name: gb_est75:BE794311

```

seq_documentation_block:
LOCUS      BE794311          688 bp      mRNA          20-SEP-2000
DEFINITION 601591442F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3945538 5',
            mRNA sequence.
ACCESSION  BE794311
VERSION    BE794311.1  GI:10215496
KEYWORDS   EST.
SOURCE     human.

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 688)
AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue procurement: DCTD/DTp
CDNA Library preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM804 row: c column: 11
High quality sequence stop: 688.

FEATURES
source

location/Qualifiers
1..688
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3945538"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="MGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 169 a 190 c 185 g 144 t
ORIGIN

alignment_scores:

Quality: 1131.00 Length: 230
Ratio: 4.982 Gaps: 1
Percent Similarity: 98.696 Percent Identity: 98.696

alignment_block:

US-09-434-382-2 x BE794311

Align seg 1/1 to: BE794311 from: 1 to: 688

```

258  LysAlaLysGluMetGlyLeuProValGlyThrAlaAlaIleAlaProII 274
|||||
3  AAAGCAAAGGAGATGGCCCTCCAGTTGGGACAGCTGCCATCGCTCCAT 52
|||||
274  eIleAlaAlaValLysAspGlyLysSerIleThrHisGluGlyArgGluI 291
|||||
53  CATTCGTGCTGTCAAGGACGGGAAAGCATCATCATCAAGGAAGAGAGA 102
|||||
291  leLeuAlaGluLeuLeuCysThrProProAspProGlyAlaAlaPheVal 307
|||||
103  TTTTGGCTGAAGAGCTGTACTCTCCAGATCCTGGTGTCTTTGTG 152
|||||
308  ValValGluCysProAspGluSerPheIleGlnProIleCysGluAsnAl 324
|||||
153  GTGTAAGTGTCCAGATGAAGCTTCAATCAACCCATCTGTGAGATGC 202
|||||
324  aThrPheGlnArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValV 341
|||||
203  CACCTTTTCAGAGGTACCAAGGAAGGAGAGATGCCCGCTGGCTTGTGG 252
|||||
341  alHisMetAlaProAlaSerValLeuValAspSerArgTyrGlnGlnTrp 357
|||||
252  TTCATGTGGCCCCAGCATCTGTCTTGTGGACAGGATACCCAGCATGG 302

```

```

358 MetGluArgPheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCy 374
|||||
303 ATGGAGAGTTTGGGCTTGCACACCCAGCACTTGGCTCTGAATGAGAACTG 352
|||||
374 sAlaSerValHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnL 391
|||||
353 TGCCCTCAGTTCAACACCTTCGAGGACCAAGATTCAAACCCAGCTCAACC 402
|||||
391 euIleHisProAspIlePheProLeuLeuThrSerPheArgCysLysLys 407
|||||
403 TCATCCACCGGACATCTCCCTGCTCACCAGTTTCGGCTGTAGAAG 452
|||||
408 GluGlyProThrLeuSerValProMetValGlnGlyGluCysLeuLeuL 424
|||||
453 GAGGGCCCCACCTC. AGTGTGCCATGTTTCCAGGTGAATGCTCCTCAA 501
|||||
424 sTyrglnLeuArgProArgArgGluTrpGlnArgAspAlaIleIleThrC 441
|||||
502 GTACCAGCTCCGTCGAGGAGGAGTGCAGAGGATGCCATTATTACTT 551
|||||
441 yAsnProGluGluPheIleValGluAlaLeuGlnLeuProAsnPheGln 457
|||||
552 GCAATCCTGAGGAATTCATAGTTGAGGCGCTGCAGCTTCCCAACTTCCAG 601
|||||
458 GlnSerValGlnGluTyArgArgSerAlaGlnAspGlyProAlaProAl 474
|||||
602 CAGAGCGTGACAGGAGTACAGGAGGAGTGCAGGACGG. CCAGCCCCAGC 650
|||||
474 acLysArgSerGln.TyrProGluIleIlePheLeu 486
|||||
651 AGAGACAGAGAGTCAAGTACCAGAAATCATCTTCTCTG 688
|||||

```

seq_name: gb_est77:BE892893

seq_documentation_block: 790 bp mRNA EST 29-SEP-2000
 LOCUS BE892893 601435738F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3920792 5',
 DEFINITION mRNA sequence.

ACCESSION BE892893
 VERSION BE892893.1 GI:10353525
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 790)
 AUTHORS NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC/DCTD/DRP

cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 --- Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov

Plate: LLCW739 row: 1 column: 09
 High quality sequence stop: 662.

FEATURES
 source Location/Qualifiers
 1..790
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3920792"
 /clone_lib="NIH_MGC_72"
 /tissue_type="melanotic
 /lab_host="DH10B (phage-resistant)"
 /notes="Organ: skin; Vector: pCMV-SpOPE6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."

BASE COUNT 185 a 221 c 220 g 164 t
 ORIGIN
 alignment_scores:
 Quality: 1123.00 Length: 267
 Ratio: 4.528 Gaps: 10
 Percent Similarity: 92.884 Percent Identity: 88.764
 alignment_block:
 US-09-434-382-2 x BE892893
 Align seg 1/1 to: BE892893 from: 1 to: 790
 302 ProGlyAlaAlaPheValValGluCysProAspGluSerPheIleGl 318
 |||||
 2 CTTGTGCTGCTTTTGTGTGTAGTATGTCAGATGAAGCTTCATTCA 51
 |||||
 318 nProIleCysGluAsnAlaThrPheGlnArgTyrGlnGlyLysAlaAspA 335
 |||||
 52 ACCCATCTGTGAGATGCCACCTTTCAGAGGTACCAAGGAAGGCAGATG 101
 |||||
 335 laProValAlaLeuValValHisMetAlaProAlaSerValLeuValAsp 351
 |||||
 102 CCCCCGTGCCCTTGTGTGTTCATATGCCGCCAGCATCTGTCTGTGGAC 151
 |||||
 352 SerArgTyrGlnGlnTrpMetGluArgPheGlyProAspThrGlnHisLe 368
 |||||
 152 AGCAGGTACCCAGCAGTGGATGGAGAGTTTGGGCTGCACCCAGCACCTT 201
 |||||
 368 uValLeuAsnGluAsnCysAlaSerValHisAsnLeuArgSerHisLysI 385
 |||||
 202 GGTCTCTGAATGAGAACTGTGCCTCAGTTCACAACCTTCGCAGCCACA 251
 |||||
 385 leGlnThrGlnLeuAsnLeuIleHisProAspIlePheProLeuLeuThr 401
 |||||
 252 TTCAACCCAGCTCAACCTCATCCACCCGGACATCTTCCCTCTGCTCAC 301
 |||||
 402 SerPheArgCysLysLysGluGlyProThrLeuSerValProMet. ValG 418
 |||||
 302 AGTTTCCCTGTAAAGAGAGGGCCCCCCTCAGTGTGCCCATGGGTTTC 351
 |||||
 418 lnglyGluCysLeuLeuLysTyrGlnLeuArgProArgGluTrpGln 434
 |||||
 352 AGGGTGAATGCTCTCCTCAAGTACCAGCTCCGTCGCCAGGAGGAGTGGCAG 401
 |||||
 435 ArgAspAlaIleIleThrCysAsnProGluGluPheIleValGluAlaLe 451
 |||||
 402 AGGGATGCCATTATTACTTGAATCCTGAGGAATTCATAGTTGAGCGCT 451
 |||||
 451 uGlnLeu. ProAsnPheGlnSerValGlnGluTyrArgArgSerAla 467
 |||||
 452 GCAGTTTCCCAACTTCCAGCAGCGTGCAGGAGTACAGGAGGAGTGGC 501
 |||||
 468 GlnAspGlyProAlaProAlaGluLysArgSerGlnTyr. ProGluIleI 484
 |||||
 502 CAGGACGCCGCCAGCCAGCAGAGAGAAAGATCAGTACCCCAAGAATCA 551
 |||||
 484 lePheLeuGlyThrGlySerAlaIleProMet. LysIleArg. AsnVal 500
 |||||
 552 TCTTCTTTGGAACAGGGTCTGCCATCCCGCGAGAAAGATTCGAACATGCA 601
 |||||
 500 erAlaThrLeu...ValAsnIleSerProAspThrSerLeuLeuLeuAsp 515
 |||||
 602 GTGCCACACTTGGTCTACTATAAGCCCCGACACGCTCTCTGCTACTGGAC 651
 |||||
 516 CysGlyGluGlyThrPhe. GlyGlnLeuCysArgHisTyrGly. AspGln 531
 |||||
 652 TGTGTGTAGGGCACATTGGGGCAAGTGTGGCGCCATTACGGAAGACCCAG 701
 |||||
 532 ValAspArgValLeuGlyThr. LeuAlaAlaValPheVal...SerHisL 547
 |||||
 702 GTGGCCAGGGTCTCTGGGGACACCTGGGGCTTGTGGGTTTGGGTTCCCA 751
 |||||

547 euHisAlaAspHisThrGlyLeuProSerIle 558
 |||||
 752 GGCACGACAGATCACCAACGGCGTTGACAAAT 786

seq_name: gb_est77:BE900936

seq_documentation_block: 745 bp mRNA 29-SEP-2000
 LOCUS BE900936 601674206f1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3957240 5',
 DEFINITION mRNA sequence.

ACCESSION BE900936

VERSION BE900936.1 GI:10389609

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 745)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LCM834 row: k column: 01

High quality sequence stop: 734.

Location/Qualifiers

FEATURES

source

1..745

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3957240"

/lab_host="DH10B (phage-resistant)"

/note="Organ: placenta; Vector: pOTB7; Site: 1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 165 a 206 c 215 g 159 t

ORIGIN

alignment_scores:

Quality: 1120.00 Length: 252

Ratio: 4.590 Gaps: 7

Percent Similarity: 96.825 Percent Identity: 95.635

alignment_block:

US-09-434-382-2 x BE900936

Align seg 1/1 to: BE900936 from: 1 to: 745

509 AspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuCy 525

|||||

2 GACACGCTCTGCTACTGACTGTGGTGGGACACATTTGGGACGCTGTG 51

|||||

525 sArgHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAlaAlay 542

|||||

52 CCCTCATTTACGGAGACAGGTGCACAGGGTCTCTGGCCACCTGGCTGCTG 101

|||||

542 alpHeValSerHisLeuHisAlaAspHisThrGlyLeuProSerIle 558

|||||

|||||

102 TGTT.GTGTCCCACTGCGCAGAGATCACACACGGGCTTGCCAAGTATC 150
 559 LeuLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisPr 575
 |||||
 151 TTGCTGCAGAGAGAACGCGCTTTGGCATCTTTGGAAAGCGCTTCACCC 200
 575 oLeuLeuValValAlaProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrH 592
 |||||
 201 TTTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 250
 592 isAsnGlnCysGlnGlnValLeuHisHisIleSerMetIleProAlaLys 608
 |||||
 251 ACAACCACTGCGCAGAGGTCTGCACCAACATCATGATGATTCCTGCCAA 300
 609 CysLeuGlnGlnGlyAlaGluIleSerSerProAlaValGluArgLeuI 625
 |||||
 301 TGCCTTTCAAGGAGGCTGAGATCTCCAGTCTCCAGTGGAAAGATTGAT 350
 625 eSerSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuV 642
 |||||
 351 CAGTTCCGCTTTCGCAACATGTGATTTGGAAGAGTTTCAGACCTGTCTGG 400
 642 aArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGly 658
 |||||
 401 TGCAGCACTGCAAGCATGCTTTGGCTGTGCTGGTGCACACCTCTGGC 450
 659 TrpLysValValTyrSerGlyAspThrMetProCysGluAlaLeuValAr 675
 |||||
 451 TGGAAAGTGGTCTATTCCGG.GACACCATCCCTCGGAGGCTCTGGTCCG 499
 675 gMetGlyLysAspAlaThrLeuLeuIleHisGluAlaThr.LeuGluAsp 691
 |||||
 500 GATGGGAAAGATGCCACCTCTCTGATACATGAAGCCACCCCTGGAAGAT 549
 692 GlyLeuGluGluGluAlaValGluLysThrHisSerThrThrSerGlnAl 708
 |||||
 550 GGTGTGAAG 599
 708 alLeSerValGlyMetArgMetAsn.AlaGluPhe.IleMetLeuAsnH 724
 |||||
 600 CATCAGCTGGGATCGGGATGAACCGCGGAGTTCATTATGCTGAACCA 649
 724 sPheSerGlnArgTyrAlaLys..ValProLeuPheSerProAsnPheSe 740
 |||||
 650 ATTACGCAAGAGGTATGCCAAAGGGTCCCTCTCTTCAGCCCAAAATTCAG 699
 740 r.GluLysValGly.ValAlaPheAsp.HisMetLysValCys 753
 |||||
 700 CGGAGAAAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 742

seq_name: gb_est77:BE902696

seq_documentation_block:

LOCUS BE902696 735 bp mRNA 29-SEP-2000

DEFINITION 601677393f1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:3959926 5',

mRNA sequence.

ACCESSION BE902696

VERSION BE902696.1 GI:10393148

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 735)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
 Plate: LLCM841 row: j column: 23
 High quality sequence stop: 732.
 Location/Qualifiers

FEATURES

1..735
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3959926"
 /clone_lib="NIH_MGC_21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 168 a 206 c 203 g 157 t 1 others
 ORIGIN

alignment_scores:

Quality: 1109.00 Length: 246
 Ratio: 4.660 Gaps: 2
 Percent Similarity: 96.748 Percent Identity: 95.528

alignment_block:

US-09-434-382-2 x BE902696

Align seg 1/1 to: BE902696 from: 1 to: 735

398 ProLeuLeuThrSerPheArgCysLysGluGlyProThrLeuSerVa 414
 3 CCCCTGCTCACCACTTCCCGCTGTAAAGAGGGCCACCCTCAGTGT 52
 414 IProMetValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgA 431
 53 GCCCATGTTCCAGGATGAATGCTCCTCAAGTACCAGCTCCGTCCAGGA 102
 431 rgGluTrpGlnArgAspAlaIleIleThr.CysAsnProGluGluPheII 447
 103 GGGAGTGGCAGAGGATGCCATTATTTACTTNGCAATCTCGAGGAATTCA 152
 447 eValGluAlaLeuGlnLeuProAsnPheGlnGlnSerValGlnGluTyrA 464
 153 ACTTGAGCGCTGCAGCTTCCCACTCCAGCAGCGGTGAGGAGTACA 202
 464 rgArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyr 480
 203 GGAGGAGTGGCAGCAGCGCCAGCCCGCAGCAGAGAAAGATCAGTAC 252
 481 ProGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleAr 497
 253 CCAGAAATCATCTTCTTGGAAACAGGGTCTGCCATCCCGATGAAGATTTCG 302
 497 gAsnValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuL 514
 303 AAATGTCAGTCCACACTTGTCAACATAAGCCCGACACGCTCTGTCTAC 352
 514 euAspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAsp 530
 353 TGGACTGTGGTGGAGGCACATTTGGGCAGCTGTGCCGTCAATCGGAGAC 402
 531 GlnValAspArgValLeuGlyThrLeuAlaValPheValSerHisLe 547
 403 CAGTGGGACAGGGTCTGGGCACCTGGCTGTGT.TGTGCCACCT 451
 547 uHisAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluA 564

452 GCACCGCAGATACCACACCGCGCTTGCCCAAGTATCTTGTGCTGCAGAGAAC 501
 564 rGAlaLeuAlaSerLeuGlyLysProLeuHisProLeuLeuValValAla 580
 502 GCGCCTTGGCATC.TTGGGAAAGCGCTTCACCCCTTGTGCTGGTGGTGGC 550
 581 ProAsnGlnLeuLysAlaTrpLeuGlnGlnTyrHisAsnGlnCysGlnG 597
 551 CCCAACCCAGCTCAAGGCTTGCTCCAGCAGTAGTACCACACAGTGCAGGA 600
 597 uValLeuHisHisIleSerMetIleProAlaLysCysLeuGlnGluGlyA 614
 601 GGTCTGTCACCATCATGATGATTCCTGC.AAATGCCCTTCAGGAAGGG 649
 614 laGluIleSerSerProAlaValGluArgLeuIleSerSerLeuLeuArg 630
 650 CTGAGATCTCCAGTCTCGAGTGGGAAGATTGATCAGTTCGCTGTTCGA 699
 631 ThrCysAspLeu.GluGluPheGlnThrCysLeu 641
 700 ACATGTGATTTGGGAAGAGTTTTCAGAACTGTCTG 733

seq_name: gb_est70:BE383336

seq_documentation_block:

LOCUS BE383336 664 bp mRNA EST 21-JUL-2000
 DEFINITION 601298249F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3628308 5',
 mRNA sequence.

ACCESSION BE383336

VERSION BE383336.1 GI:9328701

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS 1 (Bases 1 to 664)

TITLE NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel.: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM313 row: a column: 13

High quality sequence stop: 662.

FEATURES

source

1..664
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3628308"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 Note: This is a NIH_MGC Library.

BASE COUNT

ORIGIN

154 a 199 c 174 g 137 t

alignment_scores:

Quality: 1086.50

Ratio: 5.053

Length: 221

Gaps: 1

Percent Similarity: 97.285 Percent Identity: 95.023

alignment_block:

US-09-434-382-2 x BE383336 ..

Align seg 1/1 to: BE383336 from: 1 to: 664

```

365 ThrGlnHisLeuValLeuAsnGluAsnCysAlaSerValHisAsnLeuAr 381
|||||
1 ACCACGACTTGGTCTCGAATGAGAACTGGCTCAGTTTCACAACCTCG 50
|||||
381 gSerHisLysIleGlnThrGlnLeuAsnLeuLysHisProAspIlePheP 398
|||||
51 CAGCCACAGATTCAACCCAGCTCAACCTCATCCACCGGACATCTTC 100
|||||
398 roLeuLeuThrSerPheArgCysLysLysGluGlyProThrLeuSerVal 414
|||||
101 CCTGCTCACCAGTTTCCGCTGTGAAGAAGGAGGCCGCCACCTCAGTGTG 150
|||||
415 ProMetValGlnGlyCysLeuLeuLysTyrGlnLeuArgProArgAr 431
|||||
151 CCCATGGTTCAGGGTGAATGCCCTCTCAAGTACCAAGCTCCGCCAGGAG 200
|||||
431 gGluTrpGlnArgAspAlaIleThrCysAsnProGluGluPheIleV 448
|||||
201 GGAGTGGCAGAGGAGTGCATATTACTTGCATCTGAGGAATTCATAG 250
|||||
448 aGluAlaLeuGlnLeuProAsnPheGlnSerValGlnGluTyrArg 464
|||||
251 TTGACGCGCTGAGCTTCCCACTTCCAGCAGCGCTGCAGGAGTACAGG 300
|||||
465 ArgSerAlaGlnAspGlyProAlaProAlaGluLysArgSerGlnTyrPr 481
|||||
301 AGGAGTGCAGGAGCGCCAGCCCGCAGCAGAGAGAAAGTCAAGTACCC 350
|||||
481 oGluIleIlePheLeuGlyThrGlySerAlaIleProMetLysIleArgA 498
|||||
351 AGAAATCATCTTCCCTTGGAAACAGGGTCTGCCATCCGATTCGAGATT 400
|||||
498 snValSerAlaThrLeuValAsnIleSerProAspThrSerLeuLeuLeu 514
|||||
401 ATGTCAAGTCCACACTTGTCAACATTAAGCCCGACACGCTCTGTCTACTG 450
|||||
515 AspCysGlyGluGlyThrPheGlyGlnLeuCysArgHisTyrGlyAspG 531
|||||
451 GACTGTGGTGAGGACATTTGGGACGCTGTGCCGTCAATACGAGACCA 500
|||||
531 nValAspArgValLeuGlyThrLeuAlaAlaValPheValSerHisLeuH 548
|||||
501 GGTGGACAGGGTCTCGGGCACCCCTGGCTGTGTGTGTGTCACCACTGC 550
|||||
548 isAlaAspHisHisThrGlyLeuProSerIleLeuLeuGlnArgGluArg 564
|||||
551 ACGCAGATCACCACAGCGGCTTGCACAGTATCTTGTGCAAGAGAGAACG 600
|||||
565 AlaLeuAlaSerLeuGlyLysProLeuHisProLeuValValAlaPr 581
|||||
601 CGCCTGGCATCTTGGGAAAGCGCTTACCCCTTGTGCTGTGTGTG...CC 647
|||||
581 oAsnGlnLeuLys 585
|||||
648 CCAACAGTCAAG 660

```

seq_name: gb_est70:BE382353

seq_documentation_block:

```

LOCUS BE382353 692 bp mRNA EST 21-JUL-2000
DEFINITION 601298656F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3629028 5',
mRNA sequence.
ACCESSION BE382353
VERSION BE382353.1 GI:9327718
KEYWORDS EST.
SOURCE human.

```

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 692)

NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: L10M314 row: O column: 13

High quality sequence stop: 600.

FEATURES

source

Location/Qualifiers
 1..692
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3629028"
 /clone_lib="NIH_MGC_19"
 /tissue_type="neuroblastoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G).. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."

BASE COUNT

152 a 193 c 197 g 150 t

ORIGIN

alignment_scores:

Quality: 1066.00 Length: 224
 Ratio: 4.912 Gaps: 3
 Percent Similarity: 96.875 Percent Identity: 95.089

alignment_block:

US-09-434-382-2 x BE382353 ..

Align seg 1/1 to: BE382353 from: 1 to: 692

```

493 ProMetLysIleArgAsnValSerAlaThrLeuValAsnIleSerProAs 509
|||||
3 CCGATGAAGATTGCAATGTCAGTGGCACACATTTGTCAACATAAGCCCCGA 52
|||||
509 pThrSerLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuCysA 526
|||||
53 CAGCTCTCTGTACTGGACTGTGTGAGGCGACATTTGGGCGAGCTGTGCC 102
|||||
526 rGHisTyrGlyAspGlnValAspArgValLeuGlyThrLeuAlaAlaVal 542
|||||
103 GTCATTTACGGAGACAGGTCGACAGGGTCTTGGCACCCCTGGCTGTGTG 152
|||||
543 PheValSerHisLeuHisAlaAspHisHisThrGlyLeuProSerIleLe 559
|||||
153 TTTTGTGTCCACCTGCAGCGAGATCACCACACGGGCTTGCAAGTATCTT 202
|||||
559 uLeuGlnArgGluArgAlaLeuAlaSerLeuGlyLysProLeuHisProL 576
|||||
203 GCTGCGAGAGAGACGCGCTTGGCATCTTTGGGAAAGCGGCTTCACCCCT 252
|||||
576 euLeuValValAlaProAsnGlnLeuLysAlaTrpLeuGlnTyrHis 592
|||||
253 TGCTGGTGTGTGCCCCCAACACAGCTCAAGCCCTGGCTCCAGCAGTACCA 302
|||||
593 AsnGlnCysGlnGluValLeuHisHisIleSerMetIleProAlaLysCy 609

```

```

|||||
303 AACCAAGTCCAGAGGCTCCTGCACACATCAGTATTCCTGCCAAATG 352
609 sLeuGlnGluGlyAlaGluIleSerSerProAlaValGluArgLeuIleS 626
|||||
353 CCTTCAGGAAGGGCTGAGATCTCCAGTCTCGAGTGGAAAGATGATCA 402
626 erSerLeuLeuArgThrCysAspLeuGluGluPheGlnThrCysLeuVal 642
403 GTTCGCTGTCGAACATGTGATTTGGAAGAGTTTCAGACCTGTCTGCTG 452
643 ArgHisCysLysHisAlaPheGlyCysAlaLeuValHisThrSerGlyTr 659
453 CGGCACCTCAAGACATGCTTTGGCTGCTGCTGCTGCTGCTGCTGCTG 502
659 pLysValValTyrSerGlyAspThrMetProCysGluAlaLeuValArgM 676
503 GAAAGTGTCTATTCCGGGGACACCATGCTCGGAGGCTCTGTGTCGGA 552
676 etGlyLysAspAlaThrLeuLeuIleHisGluAlaThrLeuGluAspGly 692
553 TGGGAAAGATGCCACCTCTCTGATACATGAAGCCACCTG.GAAGATGCT 601
693 LeuGluGluAlaValAluGluLysThrHisSerThr.ThrSerGlnAlaI 709
602 TTGGCAGAGGAGCGCTGGGAAGA...CACACGACACAGTCCCAACCA. 647
709 leSerValGlyMetArgMet 715
648 ..TCACGTGGGATCGGGATA 665

```

seq_name: gb_est69:BE260626

```

seq_documentation_block: 938 bp mRNA EST 13-JUL-2000
LOCUS BE260626 601146116F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3161691 5',
DEFINITION mRNA sequence.
ACCESSION BE260626
VERSION BE260626.1 GI:9132065
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 938)
NIH-MGC http://www.ncbi.nlm.nih.gov/MGC/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at: image.llnl.gov
Plate: L1CM119 row: o column: 04
High quality sequence stop: 621.
Location/Qualifiers
1..938
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3161691"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGACACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit

```

FEATURES
source

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT 190 a 263 c 283 g 202 t

ORIGIN

alignment_scores:
Quality: 1005.00 Length: 290
Ratio: 4.102 Gaps: 10
Percent Similarity: 84.483 Percent Identity: 75.862

alignment_block:
US-09-434-382-2 x BE260626 ..

Align seg 1/1 to: BE260626 from: 1 to: 938

```

261 GluMetGlyLeuProValGlyThrAlaAlaIleAlaProIleIleAlaAl 277
|||||
1 GAGATGCTTTTCCCGCTCCTTGACAGCAGCAATGATGGAGCGATGGCAGC 50
277 aValLysAspGlyLysSerIleThrHisGluGlyArgGluIleLeuAlaG 294
|||||
51 TGTCCCAACTGGGAGAGCATCACTCATGAAGAGAGAGATTTGGCTG 100
294 luGluLeuCysThrProProAspProGlyAlaAlaPheValValGlu 310
|||||
101 AAGAGCTGTACTCTCCAGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 149
311 CysProAspGluSerPheIleGlnProIleCysGluAsnAlaThrPheGl 327
|||||
150 TGTCCAGATGAAGCTTCATTCAACCCCATCTGTGAGAAATGCCACCTTCA 199
327 nArgTyrGlnGlyLysAlaAspAlaProValAlaLeuValHisMetA 344
200 GAGGTACCAAGAAAGGAGATGCCCGCTGGCTGTGGTGTTCACATGG 249
344 laProAlaSerValLeuValAspSerArgTyrGlnGlnTrpMetGluArg 360
250 CCCCAGCATCTGTCTTGTGGACAGCAGGTACCAGCAGTGGATGGAGAGG 299
361 PheGlyProAspThrGlnHisLeuValLeuAsnGluAsnCysAlaSerVa 377
|||||
300 TTTGGGCTGTACACCCAGCATCTGTGCTGCTGAATGAGAACTGTGCCCTCAGT 349
377 lHisAsnLeuArgSerHisLysIleGlnThrGlnLeuAsnLeuIleHisP 394
350 TCACAACTTCGCAGCCACAGATTCACACCCAGCTCAACCTCATCCACC 399
394 roAspIlePheProLeuLeuThrSerPheArgCysLysLysGluGlyPro 410
400 CGGACATCTTCCCGCTGTCCACAGTTTCCGCTGTGAAGAGAGAGGCCCC 449
411 ThrLeuSerValProMetValGlnGlyGluCysLeuLeuLysTyrGlnLe 427
450 ACCCTCAGTGTGCCATGGTTTCAGGGTGAATGCCCTCAAGTACCAGCT 499
427 uArgProArgArgGluTrpGlnArgAspAlaIleIleThrCysAsnProG 444
500 CCGTCCAGAGGAGTGGCAGAGGGATGCCATTATTACTTGCATCTCTG 549
444 luGluPheIleValAlaLeuGlnLeu.ProAsnPheGlnGlnSerVa 460
550 AGGAATTCATGTTGAGGCGCTGAGCTTCCCACTTCCAGCAGAGCGGT 599
460 lGlnGluTyrArgSerAlaGlnAspGlyProAlaProAlaGluLysA 477
|||||
600 CGCGAGT...ACAGAGAGAGTGGCGGAGGGCCAGGCCCGCGAGAA...A 643
477 rgSerGlnTyrProGluIlePheLeuGlyThrGlySerAlaIlePro 493
644 GAAGTCGTTTCCCGGAT...CTCTCTCTG...GACGGGTCTGGCATCCCG 687
494 MetLysIleArgAsnValSerAlaThrLeuValAsnIleSerPro..... 508

```

```

688 GTGAAGATTGCA.....ATGTCGGTCCCCCGTTCCTCA 719
509 .AspThrSerLeuLeuLeuAspCysGlyGluGlyThrPheGlyGlnLeuC 525
720 TAAAGCCGCGGCTCTGTTACTACTGTTGGGTGGGCTTTGGGCG.....T 763
525 ysArgHis....TyrGlyAspGlnValAspArgValLeu.....Gly 537
764 GCCGTACGACGAGCGGGGGGCGCTGGGCCCGCTGTGTTTGTCCCGCGGG 813
538 ThrLeuAlaAlaValPhe 543
814 ACAACGGGGCGGCTTTT 831

```

seq_name: gb_est75:BE742908

seq_documentation_block: 677 bp mRNA EST 15-SEP-2000
 LOCUS BE742908 601574609f1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3835658 5',
 DEFINITION mRNA sequence.

ACCESSION BE742908
 VERSION BE742908.1 GI:10156900
 KEYWORDS EST

SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 677)

AUTHORS NIH-MGC <http://www.ncbi.nlm.nih.gov/MGC/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: DCTD/DTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM518 row: a column: 03

High quality sequence stop: 672.

FEATURES

source

```

1..677
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /clone="IMAGE:3835658"
  /clone_lib="NIH_MGC_9"
  /tissue_type="adenocarcinoma cell line"
  /lab_host="DH10B (phage-resistant)"
  /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
  EcoRI; cDNA made by oligo-dT priming. Directionally
  cloned into EcoRI/XhoI sites using the following 5'-
  adaptor: GGCAAGAG(G). Size-selected >500bp for average
  insert size 1.8kb. Library constructed by Ling Hong in
  the laboratory of Gerald M. Rubin (University of
  California, Berkeley) using ZAP-cDNA synthesis kit
  (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT

ORIGIN

alignment_scores:

```

Quality: 1000.50      Length: 226
Ratio: 4.611          Gaps: 3
Percent Similarity: 96.018      Percent Identity: 95.133

```

alignment_block:

US-09-434-382-2 x BE742908 ..

Align seg 1/1 to: BE742908 from: 1 to: 677

```

603 SerMetIleProAlaLysCysLeuGlnGluGlyAlaGluIleSerSerPr 619
1 AGTATGATTCTCTGCCAAATGCCTTCAGGAAGGGGCTGAGATCTCCAGTCC 50
619 oAlaValGluArgLeuIleSerSerLeuLeuArgThrCysAspLeuGluG 636
51 TCAGTGGAAAGATTGATCAGTTTCGCTGTTTCGGAACATGTGATTGGAAG 100
636 luPheGlnThrCysLeuValArgHisCysLysHisAlaPheGlyCysAla 652
101 AGTTTCAGACCTGTCTGTGGTGGGCACTCAAGCATGCGTTTGGCTGTGCG 150
653 LeuValHisThrSerGlyTrpLysValValTyrSerGlyAspThrMetPr 669
151 CTGGTGCACACCTCTGGCTGGAAAGTGGTCTATTTCGGGGGACACCATGC 200
669 oCysGluAlaLeuValArgMetGlyLysAspAlaThrLeuLeuIleHisG 686
201 CTGGCAGGCTCTGGTCCGGATGGGAAAGATGCCACCTCCTCATACATG 250
686 luAlaThrLeuGluAspGlyLeuGluGluAlaValGluLysThrHis 702
251 AAGCCACCTCTGGAAGATGGTTG.GAAGAGAGACAGTGGGAAAGACACAC 299
703 SerThrThrSerGlnAlaIleSerValGlyMetArgMetAsnAlaGluPh 719
300 AGCACAACTCTCCCAAGCCATCAGCGTGGGATGCGGATGAACGCGGAGTT 349
719 erIleMetLeuAsnHisPheSerGlnArgTyrAlaLysValProLeuPheS 736
350 CATTATGCTGAACCACTTCAGCCAGCGCTATGCCAAGGTCCCTCTCTTCA 399
736 erProAsnPheSerGluLysValGlyValAlaPheAspHisMetLysVal 752
400 GCCCAACTTCAGCGAGAAGTGGGAGTTGCCCTTGACCACATGAAGGTC 449
753 CysPheGlyAspPheProThrMetProLysLeuIleProProLeuLysAl 769
450 TGCTT.GGAGACTTTCACCAATGCCCAAGCTGATTCCTCCCTCAAGGC 498
769 a.LeuPheAlaGlyAspIleGluGluMetClnuArgArgGluLysArg 785
499 CCGTGTGTTGCTGGCGACATCGAGGAGATGGAGAGCGCCAGGGAAGCGG 548
786 GluLeuArgGlnValArgAlaAlaLeuLeuSerArgGluLeuAlaGly.. 801
549 GAGCTGCGGCGAGGTGCGGGGGGCC.....TCCTGTCCAGGAGCTGGCAG 592
802 .GlyLeuGluAspGlyGluProGlnLysArgAlaHisThrGluGluP 818
593 GCGGCTGGAGGATGGGAGCCTCAGCAGAAGCGGGC.CACACAGAGGAGC 641
818 roGlnAlaLysLysValArgAlaGln 826
642 CACAGCCCAAGAAGGTACAGAGCCCGAG 667

```

